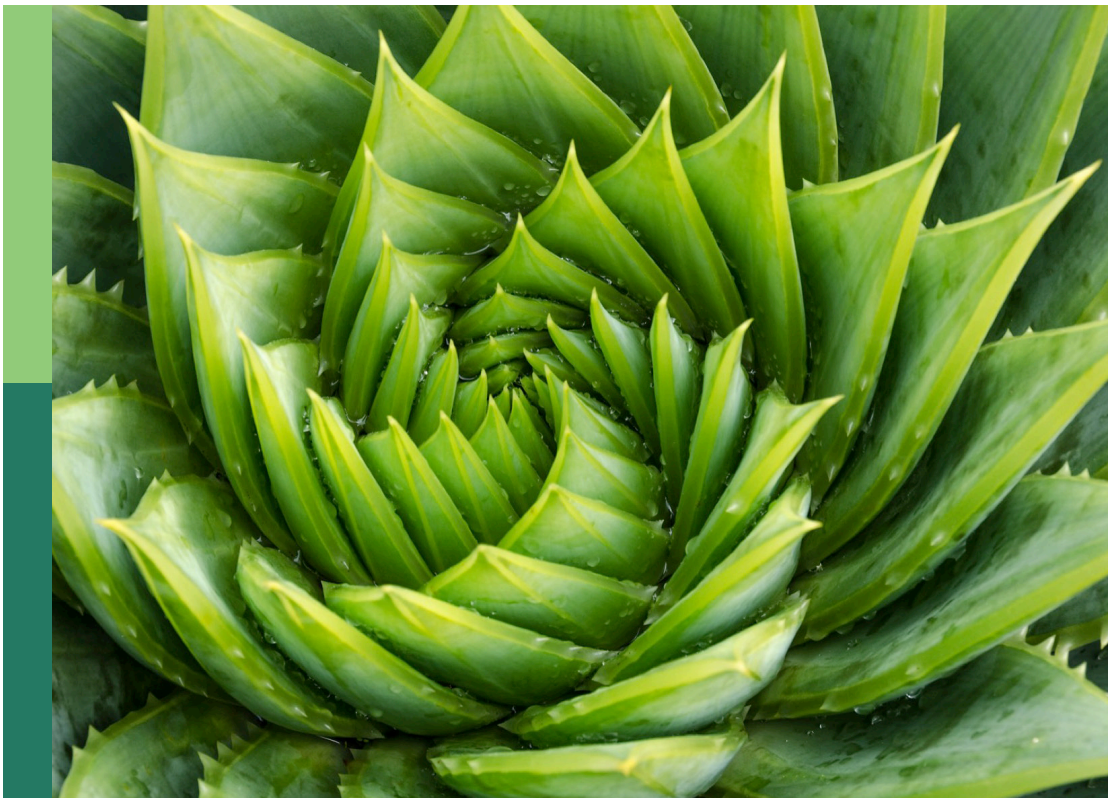
frontiers | Research Topics

Machine vision and machine learning for plant phenotyping and precision agriculture

Edited by
Huajian Liu and Zhanyou Xu

Published in
Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-4293-4
DOI 10.3389/978-2-8325-4293-4

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Machine vision and machine learning for plant phenotyping and precision agriculture

Topic editors

Huajian Liu — University of Adelaide, Australia

Zhanyou Xu — Agricultural Research Service, United States Department of Agriculture, United States

Citation

Liu, H., Xu, Z., eds. (2024). *Machine vision and machine learning for plant phenotyping and precision agriculture*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4293-4

Table of contents

06 Editorial: Machine vision and machine learning for plant phenotyping and precision agriculture
Huajian Liu and Zhanyou Xu

10 Identification of plant leaf diseases by deep learning based on channel attention and channel pruning
Riyao Chen, Haixia Qi, Yu Liang and Mingchao Yang

25 A non-destructive testing method for early detection of ginseng root diseases using machine learning technologies based on leaf hyperspectral reflectance
Guiping Zhao, Yifei Pei, Ruoqi Yang, Li Xiang, Zihan Fang, Ye Wang, Dou Yin, Jie Wu, Dan Gao, Dade Yu and Xiwen Li

40 Hyperspectral imaging with machine learning for non-destructive classification of *Astragalus membranaceus* var. *mongholicus*, *Astragalus membranaceus*, and similar seeds
Yanan Xu, Weifeng Wu, Yi Chen, Tingting Zhang, Keling Tu, Yun Hao, Hailu Cao, Xuehui Dong and Qun Sun

54 Inter-row information recognition of maize in the middle and late stages *via* LiDAR supplementary vision
Zhiqiang Li, Dongbo Xie, Lichao Liu, Hai Wang and Liqing Chen

68 Estimation of wheat tiller density using remote sensing data and machine learning methods
Jinkang Hu, Bing Zhang, Dailiang Peng, Ruyi Yu, Yao Liu, Chenchao Xiao, Cunjun Li, Tao Dong, Moren Fang, Huichun Ye, Wenjiang Huang, Binbin Lin, Mengmeng Wang, Enhui Cheng and Songlin Yang

82 Detection and counting of pigment glands in cotton leaves using improved U-Net
Lixuan She, Nan Wang, Yaxuan Xu, Guoning Wang and Limin Shao

96 Real-time determination of flowering period for field wheat based on improved YOLOv5s model
Xubin Song, Lipeng Liu, Chunying Wang, Wanteng Zhang, Yang Li, Junke Zhu, Ping Liu and Xiang Li

116 Tiller estimation method using deep neural networks
Rikuya Kinose, Yuzuko Utsumi, Masakazu Iwamura and Koichi Kise

127 A comprehensive review on acquisition of phenotypic information of Prunoideae fruits: Image technology
Xuan Liu, Na Li, Yirui Huang, Xiujun Lin and Zhenhui Ren

148 Cotton leaf segmentation with composite backbone architecture combining convolution and attention
Jingkun Yan, Tianying Yan, Weixin Ye, Xin Lv, Pan Gao and Wei Xu

160 Automatic counting of rapeseed inflorescences using deep learning method and UAV RGB imagery
Jie Li, Yi Li, Jiangwei Qiao, Li Li, Xinfa Wang, Jian Yao and Guisheng Liao

176 EmergeNet: A novel deep-learning based ensemble segmentation model for emergence timing detection of coleoptile
Aankit Das, Sruti Das Choudhury, Amit Kumar Das, Ashok Samal and Tala Awada

191 Automated extraction of pod phenotype data from micro-computed tomography
Evangeline Corcoran, Laura Siles, Smita Kurup and Sebastian Ahnert

207 A longan yield estimation approach based on UAV images and deep learning
Denghui Li, Xiaoxuan Sun, Yuhang Jia, Zhongwei Yao, Peiyi Lin, Yingyi Chen, Haobo Zhou, Zhengqi Zhou, Kaixuan Wu, Linlin Shi and Jun Li

224 Severity assessment of wheat stripe rust based on machine learning
Qian Jiang, Hongli Wang and Haiguang Wang

246 Early detection of cotton verticillium wilt based on root magnetic resonance images
Wentan Tang, Na Wu, Qinlin Xiao, Sishi Chen, Pan Gao, Yong He and Lei Feng

258 “Canopy fingerprints” for characterizing three-dimensional point cloud data of soybean canopies
Therin J. Young, Talukder Z. Jubery, Clayton N. Carley, Matthew Carroll, Soumik Sarkar, Asheesh K. Singh, Arti Singh and Baskar Ganapathysubramanian

272 Image-based phenotyping to estimate anthocyanin concentrations in lettuce
Changhyeon Kim and Marc W. van Iersel

284 An efficient tomato-detection method based on improved YOLOv4-tiny model in complex environment
Philippe Lyonel Touko Mbouembe, Guoxu Liu, Jordane Sikati, Suk Chan Kim and Jae Ho Kim

298 Self-supervised maize kernel classification and segmentation for embryo identification
David Dong, Koushik Nagasubramanian, Ruidong Wang, Ursula K. Frei, Talukder Z. Jubery, Thomas Lübberstedt and Baskar Ganapathysubramanian

310 Overlapped tobacco shred image segmentation and area computation using an improved Mask RCNN network and COT algorithm
Li Wang, Kunming Jia, Yongmin Fu, Xiaoguang Xu, Lei Fan, Qiao Wang, Wenkui Zhu and Qunfeng Niu

329 Automatic segmentation of cotton roots in high-resolution minirhizotron images based on improved OCRNet
Yuxian Huang, Jingkun Yan, Yuan Zhang, Weixin Ye, Chu Zhang, Pan Gao and Xin Lv

343 Estimation of rice seedling growth traits with an end-to-end multi-objective deep learning framework
Ziran Ye, Xiangfeng Tan, Mengdi Dai, Yue Lin, Xuting Chen, Pengcheng Nie, Yunjie Ruan and Dedong Kong

355 Development of a classification model for *Cynanchum wilfordii* and *Cynanchum auriculatum* using convolutional neural network and local interpretable model-agnostic explanation technology
Dae-Hyun Jung, Ho-Youn Kim, Jae Hee Won and Soo Hyun Park

366 A hyperspectral plant health monitoring system for space crop production
Jianwei Qin, Oscar Monje, Matthew R. Nugent, Joshua R. Finn, Aubrie E. O'Rourke, Kristine D. Wilson, Ralph F. Fritsche, Insuck Baek, Diane E. Chan and Moon S. Kim

379 Point clouds segmentation of rapeseed siliques based on sparse-dense point clouds mapping
Yuhui Qiao, Qingxi Liao, Moran Zhang, Binbin Han, Chengli Peng, Zhenhao Huang, Shaodong Wang, Guangsheng Zhou and Shengyong Xu

393 Non-destructive monitoring method for leaf area of *Brassica napus* based on image processing and deep learning
Mengcheng Li, Yitao Liao, Zhifeng Lu, Mai Sun and Hongyu Lai

406 Research on the evolutionary history of the morphological structure of cotton seeds: a new perspective based on high-resolution micro-CT technology
Yuankun Li, Guanmin Huang, Xianju Lu, Shenghao Gu, Ying Zhang, Dazhuang Li, Minkun Guo, Yongjiang Zhang and Xinyu Guo

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 27 November 2023
DOI 10.3389/fpls.2023.1331918

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Roger Deal,
Emory University, United States

*CORRESPONDENCE
Huajian Liu
huajian.liu@adelaide.edu.au

RECEIVED 02 November 2023
ACCEPTED 17 November 2023
PUBLISHED 27 November 2023

CITATION
Liu H and Xu Z (2023) Editorial: Machine vision and machine learning for plant phenotyping and precision agriculture. *Front. Plant Sci.* 14:1331918. doi: 10.3389/fpls.2023.1331918

COPYRIGHT
© 2023 Liu and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Machine vision and machine learning for plant phenotyping and precision agriculture

Huajian Liu[1]* and Zhanyou Xu[2]

[1]Australian Plant Phenomics Facility, The Plant Accelerator, School of Agriculture, Food and Wine, University of Adelaide, Urrbrae, SA, Australia, [2]United States Department of Agriculture – Agricultural Research Service - Plant Science Research, St. Paul, MN, United States

KEYWORDS

plant phenotyping, precision agriculture, machine vision, machine learning, deep learning

Editorial on the Research Topic

Machine vision and machine learning for plant phenotyping and precision agriculture

Plant phenotyping (PP) describes the physiological and biochemical properties of plants affected by both genotypes and environments. It is an emerging research field assisting the breeding and cultivation of new crop varieties to be more productive and resilient to challenging environments. Precision agriculture (PA) uses sensing technologies to observe crops and then manages them optimally to ensure that they grow in healthy conditions, have maximum productivity, and have minimal adverse effects on the environment. Traditionally, the observation of plant traits heavily relies on human experts, which is labour-intensive, time-consuming, and subjective. Although PP and PA are two different fields, they share similar sensing and data processing technologies in many respects. Recently, driven by computer and sensor technologies, machine vision (MV) and machine learning (ML) have contributed to accurate, high-throughput and nondestructive sensing and data processing technologies to PP and PA. However, these technologies are still in their infant stage, and many challenges and questions related to them still need to be addressed.

This Research Topic aims to share the latest research results on applying MV and ML to PP and PA. It demonstrates cutting-edge technologies, bottle-necks and future research directions for MV and ML in crop breeding, crop cultivation, and disease or pest management. This Research Topic of Frontiers in Plant Sciences published a total of 28 peer-reviewed research articles, including one review paper for the phenotyping of Prunoideae fruits (Liu et al.). These articles reveal the latest research trends regarding different crop species, data types and algorithms.

The summary of the published reports shows that cotton (*Gossypium*), canola or oilseed rape (*Brassica napus*), wheat (*Triticum*) and maize (*Z. mays*) are the most important crops for study in PP and PA (Figure 1A). Cotton stands out as the most frequently examined crop, with a total of five articles dedicated to it. Yan et al. developed a leaf segmentation method in the field environments. Tang et al. investigated early detection

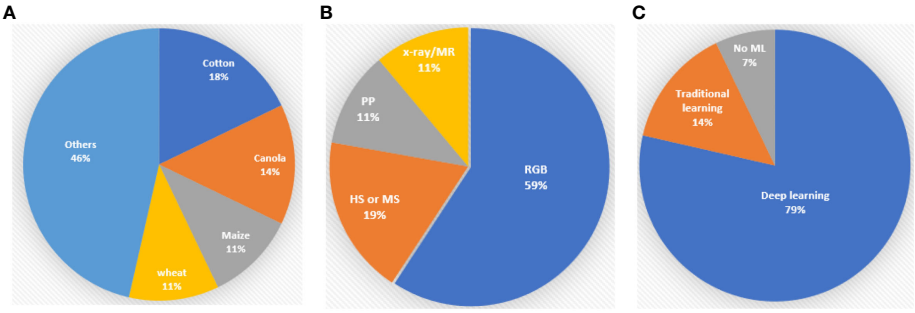


FIGURE 1
A summary of the percentage of crops, data types and machine learning algorithms in the Research Topic. PP stands for Point cloud, HS for hyperspectral data, MS for multispectral data, MR for magnetic resonance data, RGB for red-green-blue images and ML for machine learning algorithms. **(A)** Percentage of crops. **(B)** Percentage of data types. **(C)** Percentage of Machine learning algorithms.

of Verticillium wilt disease in roots. Huang et al. studied the automatic segmentation technique of roots in soil. Li et al. unveiled the evolutionary history of cotton seeds through the morphological structure of the seeds. Lastly, She et al. proposed a detection and counting method of pigment glands in cotton leaves. Canola, also known as oilseed rape, is the subject of four articles, each addressing distinct aspects of research. These include the automatic counting of inflorescences in the field environment using a UAV (Li et al.), phenotyping mature pods (Corcoran et al.), the segmentation of siliques of individual plants (Qiao et al.) and the measurement of leaf area in a laboratory environment (Li et al.). In the realm of wheat research, three pivotal studies have emerged. These encompass real-time determination of the flowering period for field wheat by detecting florets and spikelets (Song et al.), the estimation of wheat tiller density using remote sensing data (Hu et al.) and the assessment of wheat stripe rust disease severity by measuring individual leaves (Jiang et al.). Three noteworthy manuscripts, centred around maize research, have been published. These manuscripts encompass the identification of the inter-row environment during the middle and late stages for robotic navigation (Li et al.), emergence timing detection (Das et al.) and the segmentation and classification of maize seeds (Dong et al.). Other research in the main crops includes rice seedling growth traits detection (Ye et al.) and soybean canopy features description called canopy fingerprints (Young et al.). Studies focusing on vegetables or fruits have been conducted with a range of objectives. These include the estimation of anthocyanin concentrations in lettuce (Kim and Iersel), the early detection of plant stress of lettuce (Qin et al.), tomato detection (Mbouembe et al.) and the yield estimation of longan based on UAV images (Li et al.). An interesting avenue of research delves into the phenotyping of traditional Chinese herbs. Xu et al. developed a non-destructive classification method for *Astragalus membranaceus* var. *mongholicus*, *Astragalus membranaceus* and similar seeds; Zhao et al. studied the early detection of ginseng root diseases through leaves. Jung et al. investigated the classification method of *Cynanchum wilfordii* and *Cynanchum auriculatum*. Finally, Wang et al. proposed a technique for segmenting overlapped tobacco shred images and calculating their respective areas. In addition to the above-mentioned plants, two notable papers encompass the phenotyping of multiple plant species. One is about the tiller estimation method of grass plants (Kinose et al.) and another is about multiple plant leaf disease identification, providing valuable insights for disease management in various plants (Chen et al.).

Regarding data type, red-green-blue (RGB) images are the most commonly employed in PP and PA (Figure 1B) due to their cost-effectiveness and the availability of diverse algorithms. RGB images have proven to be particularly suitable for leaf segmentation (Yan et al.), tiller counting (Kinose et al.) and fruit counting (Li et al.; Mbouembe et al.). Although hyperspectral or multispectral technologies have shown great advantages over RGB imaging in PP and PA (Bruning et al., 2020; Liu et al., 2020a; Xie et al., 2021), it has not been adequately investigated. Within the scope of this Research Topic, only five publications delve into hyperspectral technologies. Xu et al. used hyperspectral imaging technologies to classify *Astragalus membranaceus* var. *mongholicus*, *Astragalus membranaceus* and similar seeds which is challenging in RGB images. Zhao et al. successfully applied hyperspectral reflectance of ginseng leaves to do early detection of root diseases. To estimate anthocyanin concentrations in lettuce at the canopy scale, Kim and Iersel used a multispectral imaging technique to develop a vegetation index called normalized difference anthocyanin index (NDAI), which is based on the optical properties of anthocyanins. Qin et al. developed a hyperspectral imaging system for plant health monitoring in a controlled-environment at NASA Kennedy Space Centre. It can conduct early detection of drought stress of twelve Dragoon lettuce samples when there are no visible symptoms. Hu et al. successfully applied hyperspectral and multispectral remote sensing data to tiller density estimation and this method could be applied to plot and county scale. LiDAR or RGB-image generated 3D point clouds demonstrated the capability to detect detailed morphological features of individual plant structures of oilseed rape (Qiao et al.), canopy structures of soybean (Young et al.) and inter-row information of maize crops (Li et al.). Micro-CT technologies showed advantages in seed phenotyping (Li et al.;Corcoran et al.) and magnetic resonance was suitable for detecting root features (Tang et al.).

This Research Topic underscores the pivotal role that machine learning, particularly deep learning, plays in PP and PA (Figure 1C). Deep learning algorithms, which are founded on artificial neural networks with multiple layers, feature prominently. Among the 28

collected articles, a substantial 22 of them incorporate deep learning algorithms. Notably, You Only Look Once (YOLO), U-net, and the attention mechanism emerge as the most frequently employed techniques. YOLO is an object detection algorithm based on convolutional neural network (CNN) architecture. It was first introduced by Redmon et al. (2016) and it has evolved over time with seven versions. YOLO has become popular because of its speed and ability to detect multiple objects in real-time. Five studies in this research domain have employed YOLO, encompassing detecting rapeseed inflorescences (Li et al.), inter-row environment identification of maize fields (Li et al.), tomato detection (Mbouembe et al.), counting the florets and spikelets of wheat (Song et al.) and estimating the yield of longan fruit (Li et al.). U-Net was introduced by Ronneberger et al. (2015) and it is a CNN architecture specifically designed for image segmentation tasks in computer vision. In this Research Topic, it was used for several segmentation tasks including cotton roots system in magnetic resonance images (Tang et al.), cotton seeds in micro-CT images (Li et al.), pigment glands in cotton leaves in RGB images (She et al.), coleoptile of maize in RGB images (Das et al.) and seeds of oilseed rape in micro-CT images (Corcoran et al.). The concept of attention models originally stemmed from sequence-to-sequence learning, as proposed by Sutskever et al. (2014). Over time, it has undergone several iterations and versions, adapting and evolving to suit various applications and challenges in the field of deep learning. It allows a neural network to focus on specific parts of input data while processing it. One of the key early papers that popularized attention mechanisms in deep learning is "Attention is all you need" by Vaswani et al. (2023). In the study of Huang et al. in order to reduce the influence of the background noise of cotton roots in minirhizotron images, they integrated a global attention module into object-contextual representation net (OCRNet) (Yuan et al., 2021) to enhance the focus of the model on the root targets. Li et al. applied an attention module to U-net to improve the segmentation of leaves of oilseed rape. The review article of Liu et al. also demonstrates that YOLO with the convolutional block attention module is a hot Research Topic in PP and PA. Four studies used traditional machine learning methods and all of them for hyperspectral data processing. Xu et al. used a support vector machine algorithm to classify *Astragalus membranaceus* var. *mongholicus*, *Astragalus membranaceus* and similar seeds. Zhao et al. adopted the random forest algorithm to detect ginseng root diseases early. Qin et al. investigated a discriminant classifier for the estimation of draught stress of lettuce. Hu et al. proposed to use a gradient-boosted regression tree and random forest to estimate wheat tiller density. For details of the traditional machine learning algorithms, readers can refer to the Sciket-Learn toolbox (ScikitLearn, 2023).

In summary, this Research Topic has compiled the most recent research articles in the domains of PP and PA. It not only disseminates the latest technologies, methods, and findings in PP and PA but also sheds light on the research trajectories and future directions within these fields. The Research Topic highlights that cotton, canola or oilseed rape, wheat and maize are the most extensively studied crops in the field of PP and PA. RGB cameras are the most commonly used sensors because of the lower cost, ease of operation and algorithms available. Although hyperspectral or multispectral images contain more information than RGB images and reveal more complex traits of plants, they have not been widely used. This is because of the higher system cost and the challenges of operation, especially in the field environment (Liu et al., 2020b). Machine learning is undeniably a pivotal component of PP and PA. Currently, the prevailing research trend involves the fusion of RGB imaging with deep learning techniques. Traditional machine learning methods continue to have significant relevance, particularly in hyperspectral data processing. However, it's worth highlighting that deep learning has not yet received extensive exploration in the realm of hyperspectral data processing. Indeed, most deep learning algorithms have their origins in RGB images, and the transition to applying them to hyperspectral images can be challenging due to the differences in spatial and spectral dimensions. However, this challenge also serves as a catalyst for a promising new research direction, where innovative approaches and techniques are required to effectively adapt and apply deep learning to hyperspectral images. The bottleneck limiting the broader application of deep learning in the agricultural industry stems from insufficient training data, as well as the substantial workload and high costs associated with manual annotation, as discussed by Dong et al. (2023). Additionally, Tang et al. highlighted the challenges of data imbalance. To address these obstacles and pave the way for more extensive deep learning use in agriculture, future research directions should encompass data augmentation techniques, self-supervised learning methods, and the generation of synthetic training data.

Author contributions

HL: Conceptualization, Formal Analysis, Project administration, Writing – original draft. ZX: Formal Analysis, Project administration, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bruning, B., Berger, B., Lewis, M., Liu, H., and Garnett, T. (2020). Approaches, applications, and future directions for hyperspectral vegetation studies: An emphasis on yield-limiting factors in wheat. *Plant Phenome J.* 3 (1). doi: 10.1002/ppj2.20007

Liu, H., Bruning, B., Garnett, T., and Berger, B. (2020a). Hyperspectral imaging and 3D technologies for plant phenotyping: From satellite to close-range sensing. *Comput. Electron. Agric.* 175. doi: 10.1016/j.compag.2020.105621

Liu, H., Bruning, B., Garnett, T., and Berger, B. (2020b). The performances of hyperspectral sensors for proximal sensing of nitrogen levels in wheat. *Sensors* 20 (16), 4550. doi: 10.3390/s20164550

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, NV, USA: IEEE), pp. 779–788. doi: 10.1109/CVPR.2016.91,ISBN:1063-6919

Ronneberger, O., Fischer, P., and Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation* (Cham: Springer International Publishing), 234–241, ISBN: . doi: 10.1007/978-3-319-24574-4_28

ScikitLearn (2023) *scikit-learn: machine Learning in Python*. Available at: https://scikit-learn.org/stable/.

Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks. *arXiv.org*. doi: 10.48550/arxiv.1409.3215

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., et al. (2023). Attention is all you need. *arXiv.org*. doi: 10.48550/arxiv.1706.03762

Xie, Y., Plett, D., and Liu, H. (2021). The promise of hyperspectral imaging for the early detection of crown rot in wheat. *AgriEngineering* 3 (4), 924–941. doi: 10.3390/agriengineering3040058

Yuan, Y., Chen, X., Chen, X., and Wang, J. (2021). Segmentation rransformer: Object-contextual representations for semantic segmentation. doi: 10.48550/arxiv.1909.11065

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 10 November 2022
DOI 10.3389/fpls.2022.1023515

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Marcin Wozniak,
Silesian University of Technology, Poland
Yunchao Tang,
Zhongkai University of Agriculture and Engineering, China

*CORRESPONDENCE
Haixia Qi
qihaixia_scau@126.com

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 19 August 2022
ACCEPTED 21 October 2022
PUBLISHED 10 November 2022

CITATION
Chen R, Qi H, Liang Y and Yang M (2022) Identification of plant leaf diseases by deep learning based on channel attention and channel pruning.
Front. Plant Sci. 13:1023515.
doi: 10.3389/fpls.2022.1023515

COPYRIGHT
© 2022 Chen, Qi, Liang and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of plant leaf diseases by deep learning based on channel attention and channel pruning

Riyao Chen[1,2], Haixia Qi[1,2,3]*, Yu Liang[1,2] and Mingchao Yang[4]

[1]College of Engineering, South China Agricultural University, Guangzhou, China, [2]National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology, Guangzhou, Guangdong, China, [3]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, Guangdong, China, [4]College of Horticulture, South China Agricultural University, Guangzhou, China

Plant diseases cause significant economic losses and food security in agriculture each year, with the critical path to reducing losses being accurate identification and timely diagnosis of plant diseases. Currently, deep neural networks have been extensively applied in plant disease identification, but such approaches still suffer from low identification accuracy and numerous parameters. Hence, this paper proposes a model combining channel attention and channel pruning called CACPNET, suitable for disease identification of common species. The channel attention mechanism adopts a local cross-channel strategy without dimensionality reduction, which is inserted into a ResNet-18-based model that combines global average pooling with global max pooling to effectively improve the features' extracting ability of plant leaf diseases. Based on the model's optimum feature extraction condition, unimportant channels are removed to reduce the model's parameters and complexity *via* the L1-norm channel weight and local compression ratio. The accuracy of CACPNET on the public dataset PlantVillage reaches 99.7% and achieves 97.7% on the local peanut leaf disease dataset. Compared with the base ResNet-18 model, the floating point operations (FLOPs) decreased by 30.35%, the parameters by 57.97%, the model size by 57.85%, and the GPU RAM requirements by 8.3%. Additionally, CACPNET outperforms current models considering inference time and throughput, reaching 22.8 ms/frame and 75.5 frames/s, respectively. The results outline that CACPNET is appealing for deployment on edge devices to improve the efficiency of precision agriculture in plant disease detection.

KEYWORDS

CACPNET, deep learning, plant leaf disease, convolutional neural network, channel attention, channel pruning

Introduction

Each year about 30 percent of global crop yields are lost due to plant diseases, resulting in direct economic losses exceeding 40 billion dollars (Dong et al., 2021). More than 821 million people have suffered from food insecurity in the past five years (Krishnamurthy et al., 2020). There are many diseases responsible for a series of losses, among which leaf spot is a common disease that often occurs in crops such as rice (Harish et al., 2008), maize (Barupal et al., 2020), and peanuts (Qi et al., 2021). Therefore, accurate identification and timely diagnosis of plant diseases are significant for plant protection (Singh and Misra, 2017). In practice, disease identification depends on professionals imposing high labor costs, lack of real-time monitoring, and unprofessional misidentification, which further increase the difficulty of identifying diseases in agriculture and lead to unstable and sharp declines in yields and food security problems (Ferentinos, 2018). Thus, intelligent and accurate identification of plant diseases without relying on manpower remains challenging for the precision agriculture field (Donatelli et al., 2017).

Recent advances in computer technology afford image classification, object detection, and natural language processing using deep learning (Li et al., 2018; Sharma and Mir, 2020; Otter et al., 2021). Currently, several deep neural network (DNN) models have been developed based on CNN features for image feature extraction, e.g., AlexNet, VGG, ResNet, and DenseNet (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; He et al., 2016; Huang et al., 2017).

Due to the powerful feature extraction capability of deep learning, researchers have already applied the above models to plant disease identification (Dhaka et al., 2021). For instance, an improved AlexNet model was used on rice diseases achieving a recognition accuracy of 95.4% (Lu et al., 2017). VGG was used on cucumber diseases after improving the fully connected layer (Zhang et al., 2019). Furthermore, ResNet-50 identified the grapevine yellows symptoms (Cruz et al., 2019), GoogLeNet was applied for disease identification in maize, tomato, and eggplant (Li et al., 2020; PAN et al., 2022), while DenseNet was used to classify nutrient deficiencies in rice crop (Sathyavani et al., 2021). The studies above demonstrate that deep neural networks improve plant disease recognition accuracy but still impose an extremely high computational cost because the models have many parameters. Specifically, the VGG-16 model has 138 million parameters and requires 15.484 Giga Floating Point Operations (GFLOPs) to conduct image recognition (Simonyan and Zisserman, 2014). The above models containing a large number of parameters are not efficient to run on plant protection equipment with limited computing power such as unmanned aerial vehicle and robots. Furthermore, due to the complexity of the field environment and the similarity of plant diseases, identification errors may lead to the spread of plant diseases. Accordingly, the recognition accuracy of the above study cannot meet the requirements of precision agriculture. Based on these results, directly applying DNN models to plant disease identification may not be effective. Therefore, enhancing the neural network's feature extraction capability and compressing the DNN models have become two significant challenges for precision agriculture to apply deep learning in the plant disease identification field.

Recently, to improve the model's recognition accuracy in large-scale classification tasks, the attention module has achieved remarkable results (Woo et al., 2018; Hu et al., 2019). In agriculture, the CBAM attention module based on DenseNet was utilized for wheat stripe rust recognition, which improved the accuracy rate by 5.47% compared with a native model (Mi et al., 2020). The ResNet-50 with SENet attention module has also been used to identify vegetable diseases with 97.24% accuracy after employing transfer learning (Zhao et al., 2022). The above results reveal that the attention module can effectively improve recognition accuracy but increases the computational time to process a single image. Therefore, the attention module is inefficient and suffers from a computationally intensive and complex structure. To balance the relationship between performance and complexity, an efficient channel attention mechanism module called ECA was proposed (Wang et al., 2020). The ECA module significantly improves the model's recognition accuracy by adding only a few parameters. Indeed, the crop disease model based on the ECA module was validated on the AI Challenger 2018 dataset, PlantVillage dataset, and self-collected cucumber disease dataset, attaining recognition accuracies of 86.35%, 99.74%, and 98.54%, respectively (Gao et al., 2021). Although the attention module improved the recognition accuracy of crop diseases, it did not reduce the model's parameter redundancy in the feature extraction process. Nevertheless, models with complex structures and excessive parameters impose significant hardware resource consumption and reduce recognition efficiency. Therefore, developing a model that achieves high accuracy while being sufficiently lightweight is still a challenge in plant disease recognition.

With advances in the Internet of Things and machine vision, mobile platforms such as unmanned aerial vehicle and robots make precision agriculture develop quickly (Tang et al., 2020; Bouguettaya et al., 2022). Due to the conflict between the high computational power requirements of the models and the limited computational power of plant protection equipment, it is a challenging task to deploy plant disease detection models on mobile platforms (Neupane and Baysal-Gurel, 2021). Currently, mobile devices are mostly used as a means of image acquisition, with disease images being transferred to more capable devices for identification (Xenakis et al., 2020). Nevertheless, recent research highlights that image recognition can be achieved using shallow networks as well (Kundu et al., 2021; Wieczorek et al., 2022), with model pruning being an effective model compression method whose core strategy is reducing the DNN's complexity *via* discarding redundant and

uninformative weights (Han et al., 2015). After pruning, the model achieves an apparent acceleration while being lightweight. Adding sparse constraints in the training stage can reduce the model's number of neurons and thus reduce the parameters and memory occupation (Zhou et al., 2016). However, the recognition accuracy can be significantly reduced due to discarding important parameters (Guo et al., 2020). Hence, it should be noted that a valid channel pruning metric must reduce the impact on model accuracy and consider the channel's importance in different layers.

In the precision agriculture field, deep learning is widely used in plant disease detection, but it still faces the problems of inefficient accuracy and excessive computational cost. In addition, the recognition rate is also an issue worthy of attention while applying the model to real-time detection of plant diseases in the field. The attention mechanism can effectively improve the identification accuracy of the model, while DNN complexity increases when adding an attention module. The ECA module uses a local cross-channel interaction strategy without dimensionality reduction, which improves accuracy without bringing in a massive quantity of parameters. However, the local feature extraction ability of the ECA module is limited and thus unable to extract the features of plant diseases well. In real time detection of plant diseases, the low recognition rate of models is one of the main factors limiting their detection effectiveness. Interestingly, pruning methods can be applied to model compression to achieve model acceleration. However, model pruning may decrease model accuracy. In addition, there is a lack of highly accurate and lightweight models that can be deployed to terminal inspection equipment in plant protection. Therefore, we propose the CACPNET model, which combines channel attention and pruning to solve the above-mentioned problems. The main contributions are summarized as follows:

1. The ECA module is modified to improve the model's ability to identify diseases for plant leaf diseases.
2. Without a significant loss of the model's accuracy, the model is channel pruned based on the channel weight importance and the local compression ratio. This strategy affords a highly accurate and lightweight model.
3. Model validation is performed using the public dataset PlantVillage and our peanut leaf disease dataset. The model's performance is analyzed based on accuracy, F1 score, FLOPs, parameters cardinality, model size, and GPU RAM.
4. The model's operation is simulated on the plant protection detection equipment, and the model's recognition rate is analyzed based on inference time and throughput metrics.
5. This study fills the research gap in real-time detection of leaf diseases, including peanuts, potatoes, apples, and other 15 crops and 43 diseases. Meanwhile CACPNET can be used for training and identification of other plant diseases.

Materials and methods

Dataset acquisition

This paper utilizes two datasets for experiments, namely the PlantVillage and the peanut leaf disease dataset we collected.

PlantVillage dataset

This dataset comprises 54634 leaf images divided into 38 disease classes from 14 species: apple, blueberry, cherry, corn, grape, orange, peach, pepper, potato, raspberry, soybean, squash, strawberry, and tomato. The details on the PlantVillage dataset are presented in Table 1. The dataset is randomly divided into a training and test set according to a 4:1 ratio with a uniform image resolution of 224×224 pixels.

Peanut leaf disease dataset

This is our own collected disease dataset from peanut leaves, collected from the Agronomic experimental base of South China Agricultural University. This dataset has 6033 disease images from five categories, namely healthy leaves (HL), rust disease on a single leaf (RD), leaf-spot disease on a single leaf (LSD), scorch disease on a single leaf (SD), and both rust disease and scorch disease on a single leaf (SD+RD) (Figure 1). The above diseases are common types of diseases in peanuts, which are significant factors causing peanut yield decline.

The images are cropped, sorted, and labeled to select 300 leaves per category and divided into a training and test set according to a 4:1 ratio. We amplified the images to 7500 using the imgAug library, which applied data augmentation by rotating the images by 90, 180, and 270 degrees and employing horizontal and vertical flips (Table 2). Table 2 reports the details of the peanut disease leaves dataset. All images from this dataset are uniformly resized to 224×224 pixels before being input to the model.

Operating environment and parameter setup

All trials are implemented on a Dell Precision 3640 PC (CPU I9-10900, 32GB RAM), utilizing an Nvidia GeForce RTX 2080 Super 8GB graphics card. Considering the software, we relied on Windows 10, Python 3.8.5, and Torch 1.9.0+cu102.

TABLE 1 Basic information of the PlantVillage.

Crop	Class	Train set	Test set
Apple	Apple_scab	504	126
	Black_rot	497	124
	Cedar_apple_rust	220	55
	Healthy	1316	329
Blueberry	Healthy	1202	300
Cherry	Healthy	684	170
	Powdery_mildew	842	210
Corn	Cercospora_leaf_spot Gray_leaf_spot	411	102
	Common_rust	954	238
	Healthy	930	232
	Northern_Leaf_Blight	788	197
Grape	Black_rot	944	236
	Esca_(Black_Measles)	1107	276
	Healthy	339	84
	Leaf_blight_(Isariopsis_Leaf_Spot)	861	215
Orange	Haunglongbing_(Citrus_greening)	4406	1101
Peach	Bacterial_spot	1838	459
	Healthy	288	72
Pepper	Bacterial_spot	798	199
	Healthy	1183	295
Potato	Early_blight	800	200
	Healthy	122	30
	Late_blight	800	200
Raspberry	Healthy	297	74
Soybean	Healthy	4072	1018
Squash	Powdery_mildew	1468	367
Strawberry	Healthy	365	91
	Leaf_scorch	888	221
Tomato	Bacterial_spot	1702	425
	Early_blight	800	200
	Healthy	1273	318
	Late_blight	1528	381
	Leaf_mold	762	190
	Septoria_leaf_spot	1417	354
	Spider_mites Two-spotted_spider_mite	1341	664
	Target_Spot	1124	280
	Tomato_mosaic_virus	299	74
	Tomato_Yellow_Leaf_Curl_Virus	4286	1071

The subsequent trials utilize the VGG-16, ResNet-18, ResNet-50, and DenseNet-121 models, and the SENet, CBAM, ECA, and the improved ECA attention modules are added to ResNet-18. For a fair comparison of the models' performance, we employ the same training parameters: the optimizer is the stochastic gradient descent (SGD), batch size of 32, 0.001 weight decay, 5e-4 learning rate, and the loss function is the CrossEntropyLoss. For the PlantVillage dataset, we consider 200 epochs, and for the peanut leaf disease dataset, 400 epochs.

Workflow of the proposed method

Overview of the CACPNET approach

Figure 2A illustrates the process of the channel attention module inserted into the model. Specifically, we traverse all model layers and insert the channel attention module after each convolutional layer. Then the new model is trained to achieve a better performance effect. Accordingly, Figure 2B depicts removing unimportant channels using channel pruning on the model. We obtain the weight relation in the channel from the

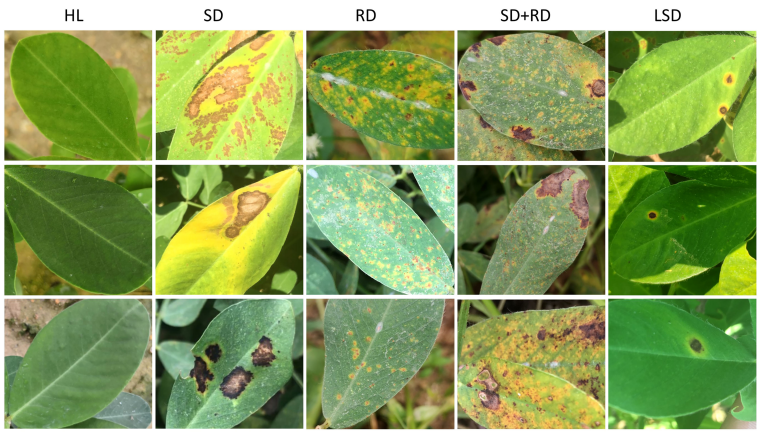


FIGURE 1
The image of peanut leaf disease. The figures show images from five categories, namely HL (healthy leaves, the first column), SD (scorch disease on a single leaf, the second column), RD (rust disease on a single leaf, the third column), SD+RD (both rust disease and scorch disease on a single leaf, the fourth column), and LSD. (leaf-spot disease on a single leaf, the fifth column).

TABLE 2 Basic information on peanut leaf disease.

Class	Train set	Test set
Healthy leaves (HL)	1200	300
Scorch disease (SD)	1200	300
Rust disease (RD)	1200	300
Both scorch and rust disease (SD+RD)	1200	300
Leaf-spot disease (LSD)	1200	300

well-trained model. Then the L1-normalization of the channel weights is calculated and ranked. The unimportant and associated channels per layer are removed based on a predetermined local compression ratio. Finally, the new model is updated with the remaining channels and retrained to achieve better performance.

Figure 2C illustrates the implementation process of CACPNET. In summary, this paper aims to develop a lightweight model with better performance and lower parameters, with the following sections introducing the details on implementing CACPNET.

Improved ECA module

This paper's channel attention module is based on an improved ECA module that uses a local cross-channel interaction strategy without dimensionality reduction. This strategy ensures that the information from the adjacent channels is correlated without losing image information, solving the correlation problem of the information contained in the plant leaf disease images. In summary, channel attention can be learned by:

$$\omega = \sigma(Wy) \tag{1}$$

which affords to gain the weight matrix (ω) if the output channels in the attention module. W is the parameter matrix ($C \times C$) y is the channel attention operation, and σ is a sigmoid activation function of the channel attention module. The parameter σ is defined as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

After the DNN's convolution operation, the output is the channel-independent parameter matrix ($C \times C$) represented as:

$$W = \begin{Bmatrix} W_1 = \begin{bmatrix} w^{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w^{c,c} \end{bmatrix} \\ \vdots \\ W_i = \begin{bmatrix} w^{1,1} & \cdots & w^{1,c} \\ \vdots & \ddots & \vdots \\ w^{c,1} & \cdots & w^{c,c} \end{bmatrix} \end{Bmatrix} \tag{3}$$

The convolution operations are performed on mutually independent channels, while the feature map information between the channels cannot interact. In order to achieve cross-channel interaction without dimensionality reduction, we use a band matrix W_k f size $k \times c$. Thus W_k can be expressed as:

$$W_k = \begin{bmatrix} w^{1,1} & \cdots & w^{1,k} & \cdots & 0 & \cdots & \cdots & w^{1,c} \\ 0 & w^{2,2} & \cdots & w^{2,k+1} & \vdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots \\ w^{c,1} & 0 & 0 & \cdots & \cdots & w^{c,c-k+1} & \cdots & w^{c,c} \end{bmatrix} \quad (4)$$

From expressions (3) and (4), we find that W contains one or more W_k. However, W_k avoids the problem of non-interacting channels between different groups in W. The range of cross-channel interactions depends on k. Nevertheless, the convolution of matrix W_k requires frequent multiplication operations, increasing the model's parameters and slowing down the running speed. Additionally, the attention mechanism operating on matrix W_k is inefficient and will slow down the model. To preserve the model's processing efficiency and to effectively obtain channel-wise feature information, we employ the ECA module, which uses a global averaging pooling (G_{avg}) operation to retain the global information of each channel. However, global average pooling may discard the disease features since plant leaf diseases occur at small locations. Instead, the global max pooling (G_{max}) can achieve translation invariance in the feature mapping and extract small location disease features (You et al., 2021). Therefore, we combine global average pooling and global max pooling affording the disease features to be effectively extracted while retaining the image's global information. The global pooling expression is as follows:

$$\begin{aligned} G(x) &= G_{avg}(x) + G_{max}(x) \\ &= \frac{1}{H \times W} \sum_{i=1,j=1}^{HW} F(x) + \max_{1 \le i \le H, 1 \le j \le W} F(x) \end{aligned} \quad (5)$$

where $F(x)$ is the aggregation feature of the input channels, and HW (e the input channel shape (Figure 2A). After global pooling, the original multidimensional input $C \times H \times W$ is transformed into a one-dimensional parameter matrix C×1×1 output. Therefore, a one-dimensional convolution kernel ($Conv1D$) can be used for efficient cross-channel information fusion operations, denoted as:

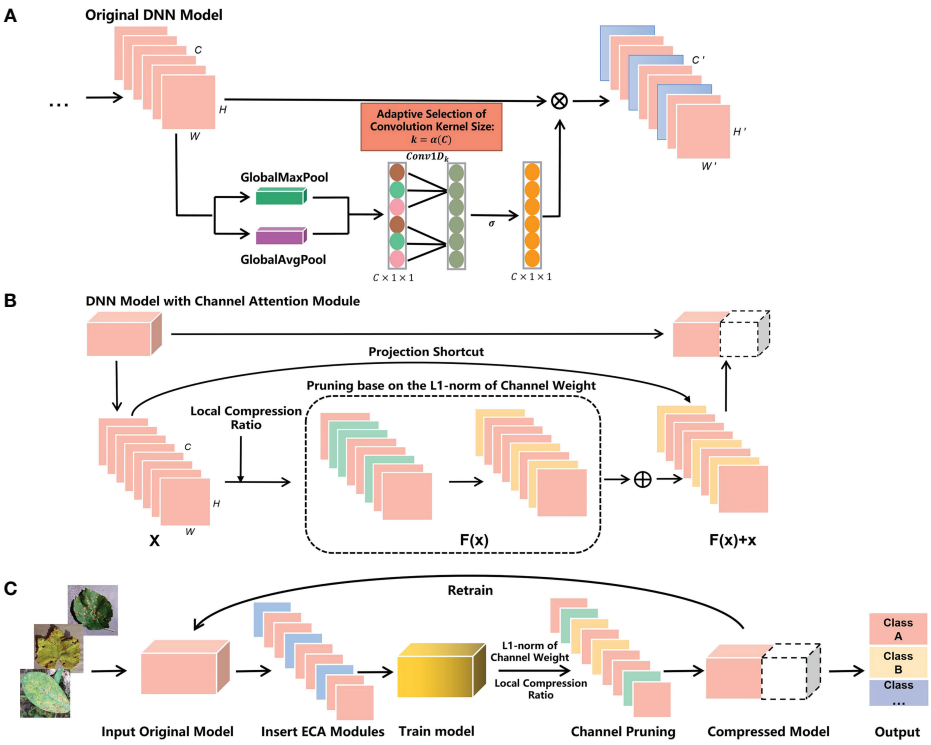


FIGURE 2
Workflow of the CACPNET. **(A)** Insert channel attention modules into the original model. The blue sections are the channel attention modules. Σ is the sigmoid activation function. **(B)** Use channel pruning to compress the DNN model. The green and yellow sections are the removed channels. **(C)** Overall structure of CACPNET.

$$\omega = \sigma(Conv1D_k(G(x))) \tag{6}$$

where k is an adjustable parameter of the one-dimensional convolution kernel that determines the scope of the cross-channel interaction. Thus, there exists a mapping relationship α between the convolution kernel size k and channel C that can be expressed as:

$$C = \alpha(k) \tag{7}$$

Channel pruning

This paper proposes a pruning method to remove unimportant channels from a well-trained model to reduce the model's parameters and complexity. In the DNN model, the channels' input and output between the layers are correlated. When the output channel of the upper layer is removed, the corresponding input channel of the lower layer also has to be removed. Therefore, the network structure of the model hierarchy must be appropriately built so that the input and output dimensions are consistent between the layers. The convolutional layer weight matrix of the DNN model consists of the input channels C_{in} the output channels C_{out} and the convolutional kernel size $H\times W$. In this paper, the weight of the output channels is calculated and sorted to judge the channels' importance, i.e., a pruning strategy for the L1-norm weights. It can be expressed as follows:

$$C_{out} = \sum_{i=1}^{C_{in}}\sum_{i=1}^{H}\sum_{i=1}^{W}\sqrt{\omega^2} = \ell - 1norm|\omega| \tag{8}$$

where ω is the weight matrix of the output channels, ℓ–$1norm|w|$ σ is a squared summation of the weight matrix w to open the roots. i.e., converting the multidimensional weight matrix $C_{in}\times H\times W$ into a one-dimensional weight matrix C_{out}. The channels' importance is obtained by sorting the weight matrix C_{out}. Since each model layer has a different effect on the extracted image information, the shallow channels are more sensitive to model pruning than the deeper channels. Furthermore, over-pruning the shallow channels will seriously affect the model's accuracy (Li et al., 2016). To reduce the influence of channel pruning on model accuracy, we introduce a local compression ratio R to prune different layers of the model. For different models and layers, the local compression ratio is an adjustable parameter, following the principle that the compression ratio of a shallow layer is smaller than a deeper layer. By introducing a local compression ratio, the model retains the crucial channels without significantly losing model accuracy and reduces the parameters and complexity. The removed channels are determined by applying L1-normalization on the weights and the local compression ratios. The number of removed channels is calculated as follows:

$$P = R \times len(C_{out}), R \in [0, 1), \tag{9}$$

where P is the number of channels removed in the current layer, R is the local compression ratio, and $len(C_{out})$ is the number of the output channels. Although a larger compression ratio can reduce the model parameters and complexity, it significantly reduces the model's accuracy. Considered together, the compression ratio of each layer for CACPNETD is set as R=[0, 0, 0.1, 0.1, 0.2, 0.2, 0.3, 0.3]. After removing P channels from this layer, the network structure in the DNN model is updated simultaneously, and the corresponding input channels of the next layer are removed.

It is worth noting that ResNet is a residual block with a shortcut connection. The expression for the residual block output y is:

$$y = F(x, \{W_i\}) + W_s x \tag{10}$$

In the ResNet model, the basic block, whose output comprises the convolutional layers' output and the residual block's output, contains two convolutional layers and a residual block. For the summation operation, the output dimension of the residual block must be the same as the convolutional layer. Therefore, the input $W_s x$ just fit the dimension of the residual learning function $F(x,\{W_i\})$. He et al., 2016). However, several previous studies have not pruned for the residual block. To ensure the consistency between the residual block channels and convolutional layer output channels after pruning, we adopt a pruning parameter sharing strategy to solve the problem of pruning the residual block. The pruning equation of the residual block can be expressed as:

$$P_r = O_r - O_{conv2}, \tag{11}$$

where P_r is the channel number removed by the residual block, O_r is the output channel of the original residual block, and O_{conv} is the output channel after pruning the previous convolutional layer. Finally, the weights after channel pruning are updated in the model.

Combination of channel attention and channel pruning

We select ResNet-18 as the base model and traverse all its layers except the residual block layer. Moreover, we insert the improved ECA attention module after each convolutional layer. Training the model containing attention modules improves recognition accuracy and establishes the channel relationship. Additionally, the trained model employs the channel pruning operation mentioned above, and finally, the weights in the model are updated after channel pruning. The pruned model is retrained to achieve better performance, with the specific CACPNET implementation presented in Algorithm 1.

Input:
D the dataset
M the original DNN model
A the ECA module
R the local compression ratios for each layer
Output:
$\mathcal{M}$ the ECA model after pruning
Procedure:
for each batch $D_t \in D$ **o**
end for
for each layer $l \in L$ **o**
end for
Retraining model

ALGORITHM 1. CACPNET ALGORITHM

Evaluate metrics

We evaluate the model's performance on the accuracy, F1 score, FLOPs, parameters, model size, and GPU RAM metrics. The accuracy and F1 score directly reflect the model's recognition performance, while the F1 score is the summed average of precision and recall. In addition, the FLOPs, parameters, model size, and GPU RAM represent the model's complexity and performance requirements of the running device. The model accuracy is expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

where TP is the prediction of positive classes as positive classes, TN is the prediction of negative classes as negative classes, FP is the prediction of negative classes as positive classes, and FN is the prediction of positive as negative classes.

The F1 score, precision, and recall are defined as:

$$F1 = \frac{2 \times Precision \times Reacall}{Precision + Recall} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

FLOPs stand for floating point operations, which are used to measure the model's complexity, defined as:

$$FLOPs = 2 \times k^2 \times C_{in} \times h_{out} \times w_{out} \times C_{out} \tag{16}$$

where k is the convolution kernel size, C_{in} the number of input channels, h_{out} and w_{out} are the height and width of the output channel, respectively, and C_{out} the number of output channels.

The parameters of the model are related to the size of the convolution kernel and the number of feature Maps, calculated as follows:

$$Parameters = C_{out} \times (C_{in} \times k^2 + 1) \tag{17}$$

Results

Ablation study on the improved attention module

The following experiments are on PlantVillage and peanut leaf disease datasets, while the model's training parameters are described in detail in the materials and methods section. To compare the effects of global average pooling and global max pooling, the ECA module of different pooling methods is inserted into ResNet-18 (Figure 2A). The results highlight that using both global average pooling and global max pooling methods simultaneously achieves better accuracy (Table 3). The features learned by the model can be visualized using Grad-CAM (Selvaraju et al., 2020). The corresponding heat map reveals that using a pooling combination enables the model to focus on the disease features (Figure 3). Interestingly, the pooling method does not increase the model's complexity, with the ablation results confirming that the proposed approach using the improved ECA module section is highly appealing. Therefore, global average and max pooling effectively extract the disease features (Figure 3 and Table 3). Thus, the ECA module using global average pooling and global max pooling is used in the following experiments.

TABLE 3 Comparison of different pooling methods in channel attention.

Description	Accuracy of PlantVillage	Accuracy of peanut	GFLOPs	Parameter
ResNet-18	98.2%	95.1%	1.819	11.180M
ECA (GlobalAvgPool) & ResNet-18	99.6%	97.5%	1.819	11.180M
ECA (GlobalMaxPool) & ResNet-18	99.5%	97.4%	1.819	11.180M
ECA (GlobalAvgPool & GlobalMaxPool) & ResNet-18	99.7%	98.0%	1.819	11.180M

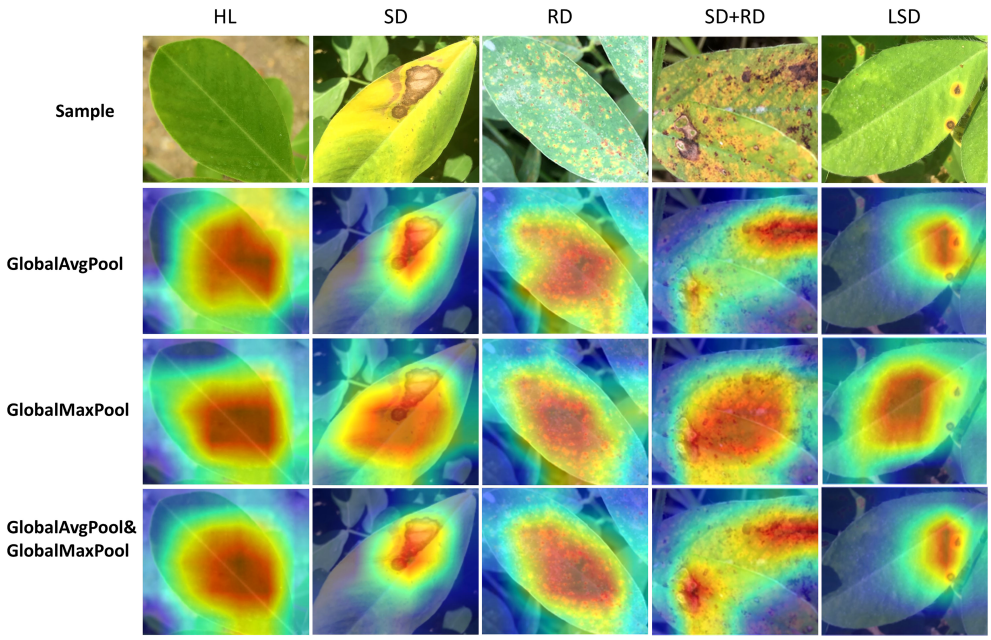


FIGURE 3
Visualized results of different pooling methods in recognizing peanut leaf disease. The dark color of the heat map represents the model's focus.

Model training and validation on the PlantVillage

This experiment challenges eight models, including the VGG-16, ResNet-18, ResNet-50, and DenseNet-121, and the remaining four are based on ResNet-18 assembled by adding SENet, CBAM, ECA, and the improved ECA attention module. All models are trained and validated using the PlantVillage dataset containing 14 species covering 38 classes of diseases (Table 1). The accuracy of all models increased along with iteration and converged at 200 epochs under the same training parameters (Figure 4A). The detailed data of each model are reported in Table 4. The accuracy of VGG-16 (97.5%) is the lowest among all models, and the parameter cardinality and FLOPs are the largest, demonstrating that a large parameter cardinality may increase the model's computational effort without effectively improving accuracy. It is worth noting that the accuracy and F1 of ResNet-50 are higher than that of ResNet-18, indicating that increasing the depth of the model improves its accuracy. However, ResNet-50 has an increased value considering Flops, parameters, model size, and GPU RAM

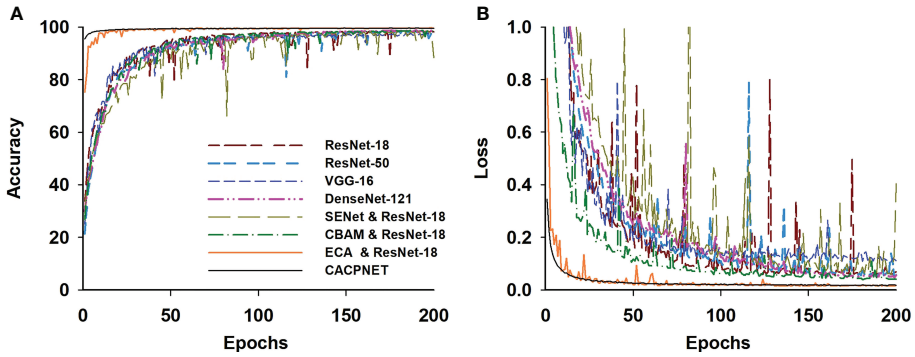


FIGURE 4
Model training and validation on the PlantVillage. **(A)** The accuracy curve of models in the PlantVillage test set; **(B)** The loss curve of models in the PlantVillage test set. Colors denote corresponding models.

TABLE 4 Indicators of model performance.

Model	Accuracy of PlantVillage	F1 score of PlantVillage	Accuracy of peanut	F1 score of peanut	GFLOPs	Parameter	Model size	GPURAM
ResNet-18	98.2%	97.4%	95.1%	95.1%	1.819	11.180M	42.7M	2.4G
ResNet-50	98.6%	97.9%	95.3%	95.3%	4.109	23.518M	90.0M	4.4G
VGG-16	97.5%	96.8%	92.4%	92.4%	15.480	134.281M	512M	7.5G
DenseNet-121	98.7%	98.1%	95.9%	95.9%	2.865	6.959M	27.1M	6.0G
SENet & ResNet-18	98.5%	97.8%	95.6%	95.6%	1.820	11.266M	43.0M	2.4G
CBAM & ResNet-18	98.7%	98.2%	95.9%	95.9%	1.821	11.267M	43.1M	2.6G
ECA & ResNet-18	99.6%	99.4%	97.5%	97.5%	1.819	11.180M	42.7M	2.4G
CACPNET	99.7%	99.5%	97.7%	97.7%	1.267	4.699M	18.0M	2.2G

requirements (Table 4). In addition, DenseNet-121 attains an appealing accuracy but requires more GPU RAM (Table 4). Regarding the attention mechanism module, the ECA module has the fastest accuracy and loss curve convergence compared to SENet and CBAM (Figure 4). The model using the ECA module convergences well on the 50th epoch. Moreover, the proposed model's accuracy reaches 99.7%, which is the best among all competitor models. Compared with the original ECA model, CACPNET presents an increased accuracy based on the improved ECA attention module with channel pruning. Moreover, CACPNET has less accuracy and loss fluctuations during the training process, suggesting that it is more robust in various plants.

Model training and validation on the peanut leaf disease dataset

When applying CACPNET on the PlantVillge dataset, it obtains better accuracy and F1 score after channel pruning and using the ECA module on 14 species (Figure 4 and Table 4). To verify the robustness of CACPNET further, the subsequent trials utilize the peanut dataset collected from an actual environment. A detailed description of the dataset acquisition is presented in the Materials and Methods section. All models converge after 400 training epochs (Figures 5A, B). However, the accuracy and F1 score of all models on the peanut test set is slightly lower than the PlantVillage (Table 4), potentially due to image interference factors originating from the actual environment. Nevertheless, the CACPNET's identification accuracy still reaches 97.7%, outperforming all competitor models. Furthermore, the accuracy and loss curves of the competitor models fluctuate more than CACPNET (Figures 4, 5A, B and Table 4), reconfirming our method's robustness.

However, rust and scorch diseases may often appear on the same leaf, easily leading to misidentification or incomplete identification. Thus, accurately identifying the two diseases when in a mixture is difficult in actual agricultural situations. However, CACPNET maintains a good identification performance, with the confusion matrix indicating a high prediction accuracy and a low error (Figure 5C). Regarding image classification identification, the ROC curves per identification class are considered a binary classification problem. The closer the curve to the upper left corner, the better the learning performance (Figure 5D). CACPNET achieves an area under the ROC curve of at least 0.98 for all five peanut leaf disease identifications (Figure 5D). In summary, CACPNET is more robust and performs better for specific class identification than the competitor methods.

Analysis of the model performance index

The previous results indicated that CACPNET performed great in both datasets considering accuracy and F1 score (Table 4). Additionally, the FLOPs, parameters, model size, and GPU RAM are essential metrics to evaluate the model's performance. FLOPs stand for floating point operations and are used to measure the complexity of a model, while the model's parameters directly determine the model's size and memory requirements. Additionally, the model's size and GPU RAM are the model's direct reflection of the required physical memory. By comparing the FLOPs and parameters among all models, we find that the FLOPs of ResNet-18 are smaller than ResNet-50, VGG-19, and DenseNet-121, while the parameters of ResNet-18 are higher than DenseNet-121 (Table 4 and Figure 6A). When adding the SENet and CBAM attention modules to ResNet-18, the FLOPs and parameter cardinality slightly increase, while FLOPs and parameters are the same when adding the ECA module (Table 4 and Figure 6A).

Although the ECA module effectively improves the model's identification accuracy without increasing its parameters and complexity (Figures 4, 5, 6A and Table 4), the parameters and complexity still limit the model from being deployed in edge devices with limited computational capabilities. Therefore, to decrease the parameters and complexity, we apply channel pruning to the model based on the improved ECA module. The FLOPs and parameters of CACPNET are significantly lower than ResNet-18 after channel pruning (Table 4 and Figure 6A). Interestingly, we maintain a high identification accuracy despite reducing the model's parameters and FLOPs through channel pruning. The model size and GPU RAM requirements of all models are presented in Figure 6B. Among all models,

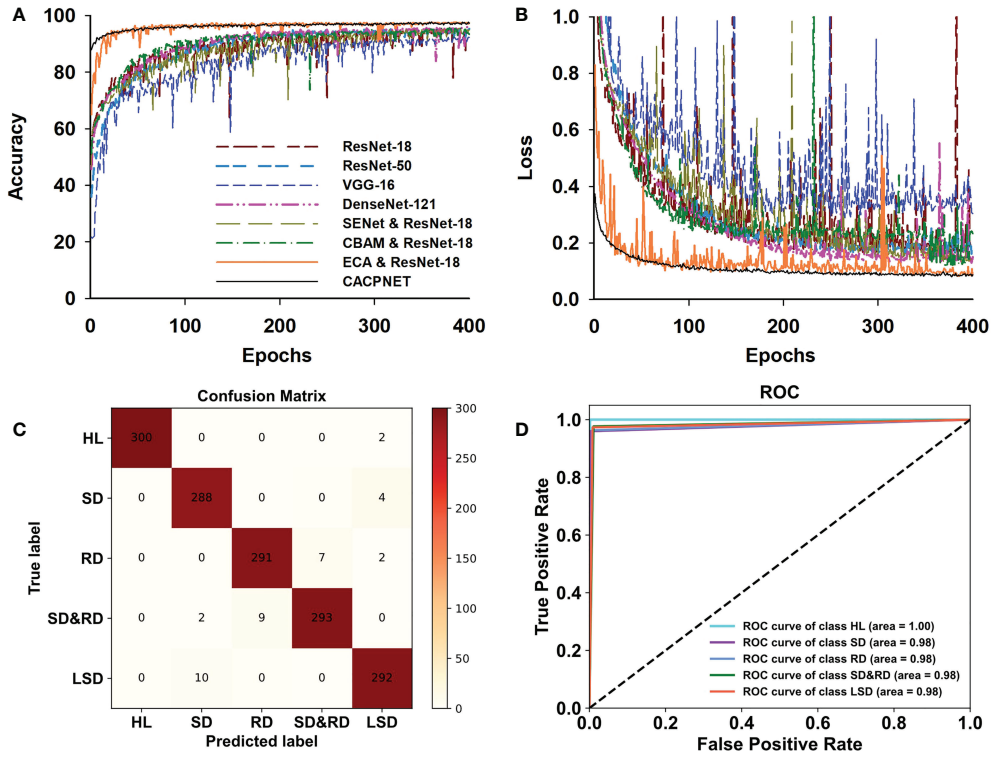


FIGURE 5
Model training and validation on the peanut leaf disease dataset. **(A)** The accuracy curve of models in the peanut test set. **(B)** the loss curve of models in the peanut test set. Colors denote corresponding models. **(C)** The confusion matrix of CACPNET in the peanut test set. The values in the corresponding column present the number of being predicted. The darker color indicates the larger quantity. **(D)** The ROC curve of CACPNET in the peanut test set. Each curve represents the identification effect of the corresponding class in the model. The closer to the upper left corner of the curve means better recognition for the corresponding class.

CACPNET has the minimum requirements on both model size and GPU RAM. Compared to ResNet-18, the FLOPs, parameters, model size, and GPU RAM of CACPNET decreased by 30.35%, 57.97%, 57.85%, and 8.3%, respectively (Table 4). In brief, without any reduction in accuracy and the F1 score, CACPNET achieves a significant reduction in FLOPs, parameters, model size, and GPU RAM. Therefore, CACPNET is more appealing according to the performance metrics.

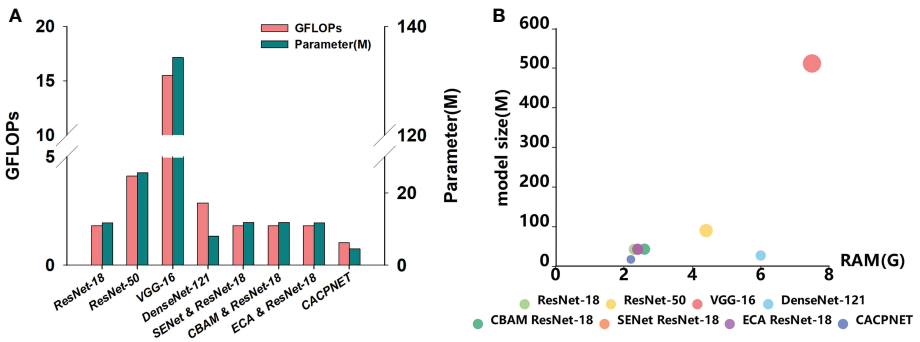


FIGURE 6
Analysis of model performance index. **(A)** Histogram of FLOPs and parameters for eight models; **(B)** Bubble chart of model size and GPU RAM for eight models.

Identification rate evaluation of the model

In the plant disease detection scenarios, the computational performance of the plant protection devices is limited, generally relying on the CPU for the computations. To validate the model's identification rate, we deployed all models for disease identification on the CPU. The metrics employed are the inference time and throughput, two important performance recognition metrics representing the time required to recognize each image and the number of images that can be processed per unit time, respectively. The experimental results highlight that the inference time and throughput of CACPNET outperform the competitor models, i.e., 22.8 ms/frame and 75.5 frames/s, respectively (Figure 7). Compared with the baseline model ResNet-18, adding the ECA, SENet, and CBAM attention modules reduces the model's recognition rate (Figure 7). However, the proposed CACPNET improves recognition accuracy and increases recognition rate (Figure 7 and Table 4). The above experimental data prove that CACPNET can operate efficiently in plant protection equipment.

Discussion

This study proposes a plant leaf disease identification model (CACPNET) that combines a channel attention mechanism and channel pruning. The PlantVillage and peanut leaf disease dataset results reveal that CACPNET's recognition accuracy and F1 score are the highest among all models (Table 4; Figures 4, 5A, B). Moreover, CACPNET has the most appealing (lowest) performance factors, such as FLOPs, parameters, model size, and GPU RAM (Table 4; Figure 6), proving that CACPNET is a lightweight model with high recognition accuracy.

Although using attention mechanisms in deep learning has made significant progress in image identification (Woo et al., 2018; Hu et al., 2019; Wang et al., 2020), its modules have not been designed for plant diseases in specific. Identification errors cannot satisfy the disease control requirements of precision agriculture. Therefore, this paper improves the ECA attention module to combine global maximum pooling with global average pooling, effectively improving the model's disease feature extraction ability.

The attention mechanism can improve the mode's recognition accuracy. However, models containing many parameters are unsuitable for planting equipment deployment with limited computing power. In particular, using the attention mechanism reduces the model's recognition and fails to meet the requirements of real-time detection regarding inference time and throughput metrics. Although the model can be simplified by channel pruning, the identification accuracy will significantly decrease due to losing important parameters (Guo et al., 2020).

Based on the existing problems presented above, we remove the unimportant channels by introducing a local compression

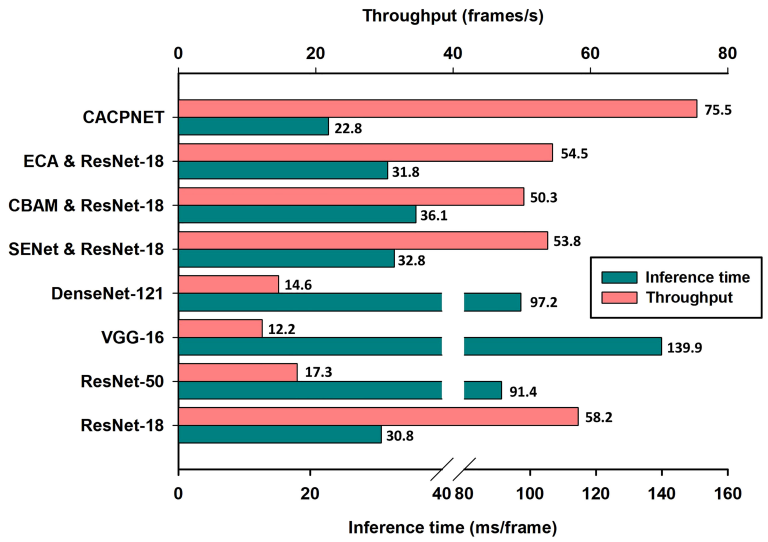


FIGURE 7
The inference time and throughput of the model. An i9-10900 was used as the CPU throughout the experiments.

ratio and an L1-norm channel weight to reduce the model's complexity and parameters. However, compared with other models, the channel-pruned CACPNET still attains the highest accuracy (Table 4; Figures 3–5). It is worth noting that the local compression ratio is a critical parameter that impacts the model's pruning effect. Although an excessive local compression ratio can significantly reduce the model's parameters and complexity, it also reduces accuracy. Therefore, it is necessary to set a reasonable local compression ratio for model pruning that depends on the model. The proposed method can maintain the model with high accuracy and lightweight at the same time, which has excellent generality to be applied to other models. In this paper, the purpose of using ResNet-18 as the base model is to obtain a more lightweight and higher accuracy model on a shallow network.

Currently, studies exist on wheat (Mi et al., 2020), apples (Yong and Ming, 2020), and grapes (Xie et al., 2020) but are constrained to a particular plant lacking universality. Unlike current methods, CACPNET is challenged on the PlantVillage database that contains 14 crops and 38 diseases, affording an identification accuracy after training of 99.7%, indicating that CACPNET can accurately identify different species and the disease characteristics of each species (Tables 1, 3, 4; Figure 4). To further verify the disease feature extraction ability of CACPNET, the leaves of peanut diseases collected in actual fields are used for training, including leaves with complex diseases such as leaves with both scorch and rust (Figure 1). Surprisingly, the identification accuracy of CACPNET reaches 97.7%, demonstrating its excellent disease feature extraction (Tables 3, 4; Figure 5).

In the precision agriculture field, real-time disease detection is an effective way to detect and control diseases in a timely manner. Inference time and throughput are important metrics to measure the real-time recognition rate of the model. Since CACPNET has a significant lead in FLOPs and parameter metrics (Figure 6 and Table 4), CACPNET is also ahead of other models in terms of inference time and throughput on CPU-based devices, with 22.8 ms/frame and 75.5 frames/s, respectively (Figure 7). The above results demonstrate that the recognition rate of CACPNET can satisfy the requirement of real-time disease detection.

Some advances in leaf disease identification have been made utilizing hyperspectral imaging (Ban et al., 2019; Nagasubramanian et al., 2019). However, the high weather or light requirements, professional operation, and extra-expensive equipment limit its overall development. Opposing, the proposed model has low image resolution requirements (224×224 or higher resolution) derived from mobile phones and pads.

Conclusion

This study proposes a lightweight model named CACPNET that is based on channel attention and channel pruning. Compared with other models, CACPNET has prominent advantages. First, CACPNET has the highest accuracy and F1 score among all competitor methods. Second, CACPNET's ability to extract plant leaf disease features can be effectively improved by combining global average pooling and global maximum pooling. In addition, CACPNET outperforms other models considering the parameters, FLOPs, Model size, and GPU RAM performance metrics. For devices relying on the CPU as the computing core, the inference time and throughput of CACPNET are superior to other models and still meet the real-time identification requirement. To sum up, CACPNET is a lightweight and highly accurate model for plant leaf disease recognition that is appropriate for lightweight model deployment in the plant protection field and promotes the development of artificial intelligence in precision agriculture. Meanwhile, this study fills the research gap in real-time detection of leaf diseases, including peanuts, potatoes, apples, and other 15 crops and 43 diseases, providing the basis for decision-making in precision agriculture.

In future work, we plan to deploy CACPNET to field robots and unmanned aerial vehicle to establish an automated disease detection platform with low inference cost. In addition, to extend CACPNET's applicability on disease identification of other plants, we will consider expanding its disease identification types through transfer learning.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://plantvillage.psu.edu/.

Author contributions

RC: Conceptualization, Methodology, Software, Data curation, Visualization Writing- original draft, review, editing and revision. HQ: conceptualization, methodology, resources, review and editing, project administration and supervision, funding acquisition. YL: data curation, validation, formal analysis, investigation. MY: visualization, editing, revision. All authors contributed to the article and approved the submitted version.

Funding

This research is funded by Characteristic innovation projects of Guangdong Provincial Department of education in 2019 (2019KTSCX015), Key-Area Research and Development Program of Guangdong Province (2019B020214005), The National Key Research and Development China Subproject (2021YFD2000701).

Acknowledgments

The authors would like to express their gratitude to EditSprings (https://www.editsprings.cn) for the expert linguistic services provided.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ban, S., Tian, M., and Chang, Q. (2019). Estimating the severity of apple mosaic disease with hyperspectral images. *Int. J. Agric. Biol. Eng.* 12 (4), 148–153. doi: 10.25165/j.ijabe.20191204.4524

Barupal, T., Meena, M., and Sharma, K. (2020). A study on preventive effects of lawsonia inermis l. bioformulations against leaf spot disease of maize. *Biocatalysis Agric. Biotechnol.* 23, 101473. doi: 10.1016/j.bcab.2019.101473

Bouguettaya, A., Zarzour, H., Kechida, A., and Taberkit, A. M. (2022). A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images. *Cluster Computing*, 1–21. doi: 10.1007/s10586-022-03627-x

Cruz, A., Ampatzidis, Y., Pierro, R., Materazzi, A., Panattoni, A., De Bellis, L, et al. (2019). Detection of grapevine yellows symptoms in vitis vinifera l. with artificial intelligence. *Comput. Electron. Agric.* 157, 63–76. doi: 10.1016/j.compag.2018.12.028

Dhaka, V. S., Meena, S. V., Rani, G., Sinwar, D., Kavita, K., Ijaz, M. F., et al. (2021). A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. *Sensors-Basel* 21, 4749. doi: 10.3390/s21144749

Donatelli, M., Magarey, R. D., Bregaglio, S., Willocquet, L., Whish, J. P., and Savary, S. (2017). Modelling the impacts of pests and diseases on agricultural systems. *Agric. Syst.* 155, 213–224. doi: 10.1016/j.agsy.2017.01.019

Dong, A., Wang, Z., Huang, J., Song, B., and Hao, G. (2021). Bioinformatic tools support decision-making in plant disease management. *Trends Plant Sci.* 26, 953–967. doi: 10.1016/j.tplants.2021.05.001

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009

Gao, R., Wang, R., Feng, L., Li, Q., and Wu, H. (2021). Dual-branch, efficient, channel attention-based crop disease identification. *Comput. Electron Agr* 190, 106410. doi: 10.1016/j.compag.2021.106410

Guo, S., Wang, Y., Li, Q., and Yan, J. (2020). Dmcp: differentiable markov channel pruning for neural networks. In *Proceedings of the Ieee Computer Society Conference On Computer Vision and Pattern Recognition*, pp. 1539–1547. Available at: https://arxiv.org/abs/2005.03354.

Han, S., Mao, H., and Dally, W. J. (2015). Deep compression: compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint* 56, 3–7.

Li, H., Kadav, A., Durdanovic, I., Samet, H., and Graf, H. P (2016). Pruning filters for efficient convnets. *arXiv preprint*.

Harish, S., Saravanakumar, D., Radjacommare, R., Ebenezar, E. G., and Seetharaman, K. (2008). Use of plant extracts and biocontrol agents for the management of brown spot disease in rice. *Biocontrol* 53, 555–567. doi: 10.1007/s10526-007-9098-9

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778. Available at: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf.

Huang, G., Liu, Z., Van Der Maaten, L., and and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4700–4708.Available at: https://arxiv.org/abs/1608.06993v5.

Hu, J., Shen, L., and Sun, G. (2019). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7132–7141. Available at: https://arxiv.org/abs/1709.01507v4.

Krishnamurthy, P. K., Choularton, R. J., and Kareiva, P. (2020). Dealing with uncertainty in famine predictions: how complex events affect food security early warning skill in the greater horn of africa. *Global Food Secur.* 26, 100374. doi: 10.1016/j.gfs.2020.100374

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM* 60 (6), 84–90. doi: 10.1145/3065386

Kundu, N., Rani, G., Dhaka, V. S., Gupta, K., Nayak, S. C., Verma, S., et al. (2021). IoT and interpretable machine learning based framework for disease prediction in pearl millet. *Sensors-Basel* 21 (16), 5386. doi: 10.3390/s21165386

Li, Z., Yang, Y., Li, Y., Guo, R., Yang, J., and Yue, J. (2020). A solanaceae disease recognition model based on se-inception. *Comput. Electron. Agric.* 178, 105792. doi: 10.1016/j.compag.2020.105792

Li, Y., Zhang, H., Xue, X., Jiang, Y., and Shen, Q. (2018). Deep learning for remote sensing image classification: a survey. *Wiley Interdisciplinary Reviews: Data Min. Knowledge Discovery* 8 (6), e1264. doi: 10.1002/widm.1264

Lu, Y., Yi, S., Zeng, N., Liu, Y., and Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. doi: 10.1016/j.neucom.2017.06.023

Mi, Z., Zhang, X., Su, J., Han, D., and Su, B. (2020). Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.558126

Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., and Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3d deep learning on hyperspectral images. *Plant Methods* 15 (1), 1–10. doi: 10.1186/s13007-019-0479-8

Neupane, K., and Baysal-Gurel, F. (2021). Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: A review. *Remote Sens-Basel* 13, 3841. doi: 10.3390/rs13193841

Otter, D. W., Medina, J. R., and Kalita, J. K. (2021). A survey of the usages of deep learning for natural language processing. *IEEE Trans. On Neural Networks Learn. Syst.* 32, 604–624. doi: 10.1109/TNNLS.2020.2979670

Pan, S. Q., Qiao, J. F., Wang, R., Yu, H. L., Wang, C., Taylor, K., et al. (2022). Intelligent diagnosis of northern corn leaf blight with deep learning model. *J. Integr. Agric.* 21 (4), 1094–1105. doi: 10.1016/S2095-3119(21)63707-3

Qi, H., Liang, Y., Ding, Q., and Zou, J. (2021). Automatic identification of peanut-leaf diseases based on stack ensemble. *Appl. Sci.* 11 (4), 1950. doi: 10.3390/app11041950

Sathyavani, R., JaganMohan, K., and Kalaavathi, B. (2021). Classification of nutrient deficiencies in rice crop using densenet-bc. *Materials Today: Proc* 56, 1783–1789. doi: 10.1016/j.matpr.2021.10.466

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: Visual explanations from deep networks *via* gradient-based localization. *Int. J. Comput. Vision* 128, 336–359. doi: 10.1007/s11263-019-01228-7

Sharma, V., and Mir, R. N. (2020). A comprehensive and systematic look up into deep learning based object detection techniques: a review. *Comput. Sci. Rev.* 38, 100301. doi: 10.1016/j.cosrev.2020.100301

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint*

Singh, V., and Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Inf. Process. Agric.* 4, 41–49. doi: 10.1016/j.inpa.2016.10.005

Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., et al. (2020). Recognition and localization methods for vision-based fruit picking robots: A review. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00510

Wang, Q., Wu, B., Zhu, P., Li, P., and Hu, Q. (2020). Eca-net: Efficient channel attention for deep convolutional neural networks. arXiv preprint.

Wieczorek, M., Siłka, J., Woźniak, M., Garg, S., and Hassan, M. M. (2022). Lightweight convolutional neural network model for human face detection in risk situations. *IEEE T Ind. Inform* 18, 4820–4829. doi: 10.1109/TII.2021.3129629

Woo, S., Park, J., Lee, J., and Kweon, I. S. (2018). Cbam: convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19. Available at: https://arxiv.org/abs/1807.06521v2.

Xenakis, A., Papastergiou, G., Gerogiannis, V. C., and Stamoulis, G. (2020). Applying a convolutional neural network in an IoT robotic system for plant disease diagnosi. In *2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA*, pp. 1–8 IEEE. doi: 10.1109/IISA50023.2020.9284356

Xie, X., Ma, Y., Liu, B., He, J., Li, S., and Wang, H. (2020). A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Front. Plant Sci.* 11, 751. doi: 10.3389/fpls.2020.00751

Yong, Z., and Ming, Z. (2020). Research on deep learning in apple leaf disease recognition. *Comput. Electron. Agric.* 168, 105146. doi: 10.1016/j.compag.2019.105146

You, H., Yu, L., Tian, S., Ma, X., Xing, Y., Xin, N., et al. (2021). Mc-net: multiple max-pooling integration module and cross multi-scale deconvolution network. *Knowledge-Based Syst.* 231, 107456. doi: 10.1016/j.knosys.2021.107456

Zhang, S., Zhang, S., Zhang, C., Wang, X., and Shi, Y. (2019). Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Comput. Electron. Agric.* 162, 422–430. doi: 10.1016/j.compag.2019.03.012

Zhao, X., Li, K., Li, Y., Ma, J., and Zhang, L. (2022). Identification method of vegetable diseases based on transfer learning and attention mechanism. *Comput. Electron Agr* 193, 106703. doi: 10.1016/j.compag.2022.106703

Zhou, H., Alvarez, J. M., and Porikli, F. (2016). "Less is more: towards compact cnns," in *European Conference on computer vision* (Cham: Springer), 662–677.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 November 2022
DOI 10.3389/fpls.2022.1031030

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Karansher Singh Sandhu,
Bayer Crop Science, United States
Jatinder Sangha,
Agriculture and Agri-Food Canada (AAFC), Canada

*CORRESPONDENCE
Xiwen Li
xwli@icmm.ac.cn
Dade Yu
ddyu@icmm.ac.cn

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 29 August 2022
ACCEPTED 02 November 2022
PUBLISHED 17 November 2022

CITATION
Zhao G, Pei Y, Yang R, Xiang L, Fang Z, Wang Y, Yin D, Wu J, Gao D, Yu D and Li X (2022) A non-destructive testing method for early detection of ginseng root diseases using machine learning technologies based on leaf hyperspectral reflectance. *Front. Plant Sci.* 13:1031030. doi: 10.3389/fpls.2022.1031030

COPYRIGHT
© 2022 Zhao, Pei, Yang, Xiang, Fang, Wang, Yin, Wu, Gao, Yu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A non-destructive testing method for early detection of ginseng root diseases using machine learning technologies based on leaf hyperspectral reflectance

Guiping Zhao[1,2], Yifei Pei[1], Ruoqi Yang[3], Li Xiang[1], Zihan Fang[4], Ye Wang[1], Dou Yin[5], Jie Wu[1,2], Dan Gao[1], Dade Yu[1*] and Xiwen Li[1*]

[1]Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China, [2]College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming, China, [3]School of Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan, China, [4]TCM Department, China National Center for Biotechnology Development, Beijing, China, [5]School of Basic Medical Sciences, Anhui Medical University, Hefei, Anhui, China

Ginseng is an important medicinal plant benefiting human health for thousands of years. Root disease is the main cause of ginseng yield loss. It is difficult to detect ginseng root disease by manual observation on the changes of leaves, as it takes a long time until symptoms appear on leaves after the infection on roots. In order to detect root diseases at early stages and limit their further spread, an efficient and non-destructive testing (NDT) method is urgently needed. Hyperspectral remote sensing technology was performed in this study to discern whether ginseng roots were diseased. Hyperspectral reflectance of leaves at 325-1,075 nm were collected from the ginsengs with no symptoms on leaves at visual. These spectra were divided into healthy and diseased groups according to the symptoms on roots after harvest. The hyperspectral data were used to construct machine learning classification models including random forest, extreme random tree (ET), adaptive boosting and gradient boosting decision tree respectively to identify diseased ginsengs, while calculating the vegetation indices and analyzing the region of specific spectral bands. The precision rates of the ET model preprocessed by savitzky golay method for the identification of healthy and diseased ginsengs reached 99% and 98%, respectively. Combined with the preliminary analysis of band importance, vegetation indices and physiological characteristics, 690-726 nm was screened out as a specific band for early detection of ginseng root diseases. Therefore, underground root diseases can be effectively detected at

an early stage by leaf hyperspectral reflectance. The NDT method for early detection of ginsengs root diseases is proposed in this study. The method is helpful in the prevention and control of root diseases of ginsengs to prevent the reduction of ginseng yield.

KEYWORDS

ginseng root diseases, hyperspectral reflectance, vegetation indices, extreme random tree, specific band

1 Introduction

Ginseng (*Panax ginseng* Mey) is one of the precious traditional herbs. The roots of ginseng are widely used as important medicinal materials for curing hypertension, stress, and neurological disorders (Ratan et al., 2021). Nowadays, wild ginsengs are endangered and cultivated ginsengs are used as main resources of ginseng products (Xu et al., 2016). The harvest rotation of cultivated ginseng is usually about 4 to 6 years (Xiao et al., 2016). Ginsengs cultivated in the same soil for a long-growth period are susceptible to root diseases (Fang et al., 2022), such as root rot, rusty root rot, red-skin, soft rot and so on (Liang et al., 2017; Lu et al., 2020). Root disease in ginseng is caused by a variety of factors, including biological factors such as pathogen infection, and abiotic factors such as soil temperature and moisture (Shang et al., 1996; Zhou et al., 2017), leading to diverse symptoms (Figure 1). The incidence rate of ginseng roots can be as high as 80% (Wang et al., 2014), which seriously reduces the production of ginseng and costs huge economic losses. So far, the commonly test method of ginseng root diseases in the field mainly relies on experiences of farmers by visually observing the symptoms of aerial parts (Farh et al., 2018), which is extremely difficult to generate and has poor accuracy, especially at the early stages of root diseases when there are no macroscopic symptoms on leaves can be observed (Lu et al., 2020).

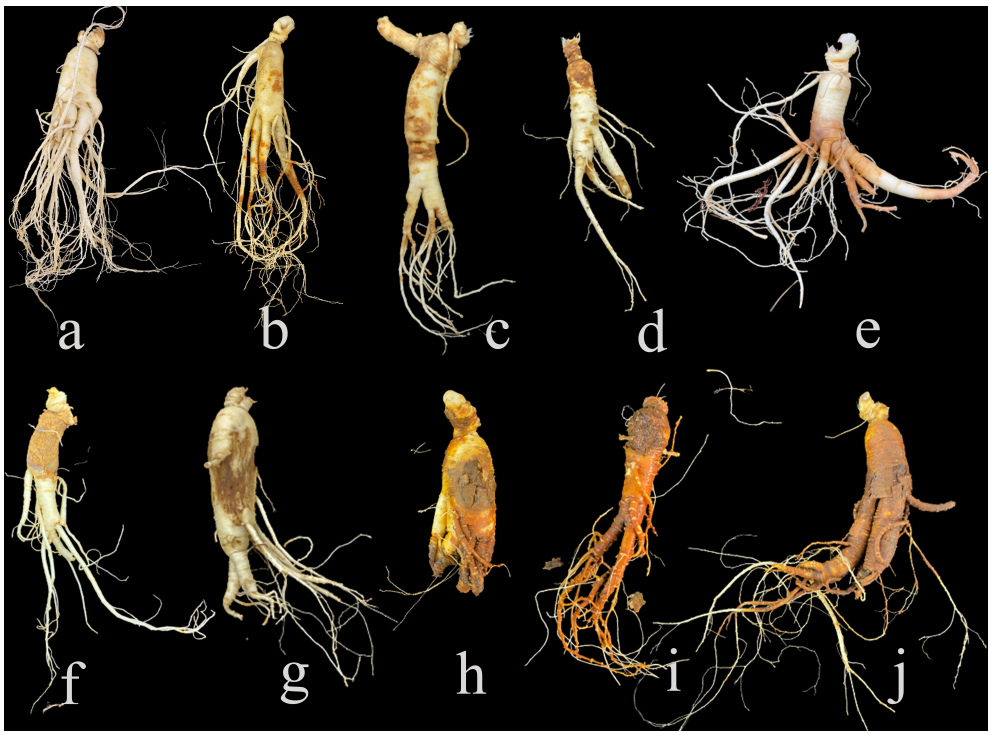


FIGURE 1
Symptoms of diseased ginseng roots collected in the field. **(A)** is healthy root, **(B–J)** are various diseased roots.

When there are visible lesions on the aboveground plant, the root has rotted and it was too late for treatment (Guan et al., 2014). Unnecessarily damages or losses are caused artificially if ginseng roots were dug out for visual detection, though it is more precise by visualization on the underground part. Thus, it is importance to detect the root diseases at their early stages, as that treatments could be taken in time.

Root diseases were often accompanied with a decrease in chlorophyll content of leaves (Fujimoto et al., 2021; Jia et al., 2021), which indicated the reduction of plant photosynthesis and external stresses (Monteoliva et al., 2021). In addition, the contents of antioxidant enzymes in leaves are also affected by root diseases (Harrach et al., 2013). The increase of enzyme activity improves the tolerance of plants to stress environments (Adavi et al., 2020; Deng et al., 2020). For example, in wheat, the antioxidant enzyme activities increased in resistance to the Bipolaris sorokiniana-common root rot (Qalavand et al., 2022). Chlorophyll content and antioxidant enzyme activity of leaves are physiological characters responding to nutrient deficiency, abiotic stresses and biotic stresses (Panda and Sarkar, 2013), the changes of which could be considered as potential indicators of root diseases of plants.

Nowadays, hyperspectral remote sensing technology has been widely used for rapid and non-destructive testing (NDT) of plant diseases (Lowe et al., 2017; Mahlein et al., 2018; Chen et al., 2022), such as wheat powdery mildew (Feng et al., 2022), fusarium head blight (Zhang et al., 2020) and decayed citrus (Zhang et al., 2021). The effective information on plant growth status is archived by screening and processing the complex and redundant hyperspectral data (Yang et al., 2020). Hyperspectral vegetation indices (VIs) were widely deployed to estimate plant biophysical and biochemical traits (Koh et al., 2022). The downy mildew severity stages in watermelon are significantly correlated with the chlorophyll green, photochemical reflectance index and normalized phaeophytinization index (Abdulridha et al., 2022). Improved accuracy of hyperspectral data processing due to advances of machine learning (ML) enabling further development of non-invasive high-throughput plant phenotyping (Mahlein et al., 2019; Arya et al., 2022). ML algorithms such as random forest (RF), support vector machine and convolutional neural networks, have improved the accuracy of hyperspectral data processing to obtain the spectral characteristics of plants during the growth period, and to finalize the early detection of plant diseases (Ghosh et al., 2022), such as the Fusarium head blight disease severity of wheat (Żelazny et al., 2021). Combined with artificial intelligence algorithms and continuous rich hyperspectral reflectance data, models could be constructed to detect and identify plant diseases (Singh et al., 2018; Singh et al., 2021). The application of ML technologies based on hyperspectral reflectance enables precise diagnosis of plant root diseases at an early stage.

To prevent the massive loss of ginseng yield caused by severe root diseases, this study proposed an NDT method using the hyperspectral remote sensing technology to detect root diseases of ginsengs at their early stages. Based on the previous field investigation, ginsengs with healthy leaves in the field of high incidence of root diseases were used. The hyperspectral reflectance data of ginseng leaves were obtained non-destructively to construct a detection model by ML. Taking the constructed model as a reference, a hyperspectral inversion model of ginseng root disease was established. We finally found that hyperspectral remote sensing technology could achieve early accurate detection of ginseng root diseases, which greatly reduced the loss of production and avoided the use of excessive pesticides.

2 Materials and methods

2.1 Experiment setup and measurements in the field

The experimental site was located at the experimental base of the Chinese Academy of Traditional Chinese Medicine in Jingyu County, Jilin Province, China (126.8°E, 42.39°N), and the ginsengs were cultivated continuously for 3-5 years. The cultivated soils were all farmland soils where ginseng had not been grown before, and the ginseng was cultivated in shade by covering with blue and yellow shade nets. The field management, including watering and fertilizing, was the same as the cultivated ginsengs in farmland as has been described by Shen et al. (2017). The measurements and harvest were conducted in 17 to 20 August 2021, a period when ginseng grows in the red-fruited stage and the average local temperature was around 15°C to 25°C. A total of 217 ginsengs with no diseased symptoms on leaves were chosen in the field for the measurements. The leaf chlorophyll contents were measured by soil plant analysis development chlorophyll meter (SPAD-502 Plus, KONICA MINOLTA, Japan) with the middle leaf of a palmately compound leaf. Three middle leaves of each ginseng were measured and each leaf was measured three times. The hyperspectral reflectance of the middle and upper parts of the largest leaf of these ginsengs were collected using a FieldSpec HandHeld 2 Spectroradiometer (HH2, Analytical Spectral Devices, Colorado, US). The wavelength of HH2 is 325-1,075 nm, and the sampling interval of HH2 is 1.4 nm. It is used in combination with a plant probe, an optical fiber and a leaf clip to avoid the disturbances of ambient light. The spectral reflectance was measured 10 times per leaf using white plate calibration in prior to measurement.

2.2 Sample collection and antioxidant enzyme measurements

After the measurement of spectral reflectance, the 217 ginseng plants were harvested. Roots with stem and leaves were dug out. They were washed with running water and thereafter dried by tissue paper. Pictures of each plant were

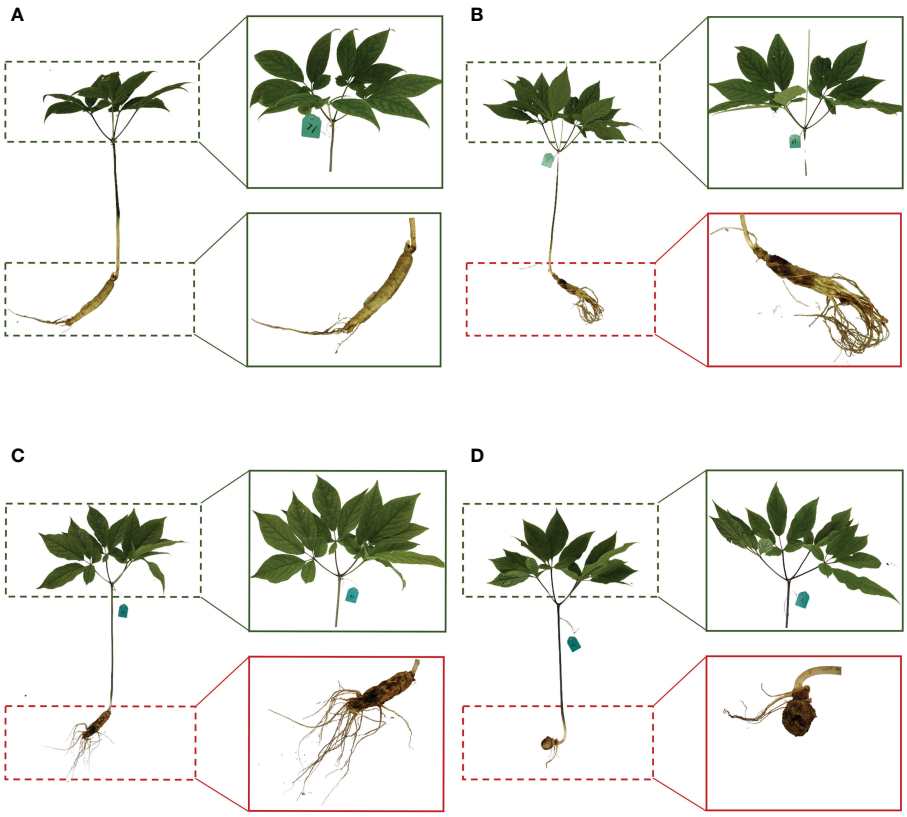


FIGURE 2
Whole plant, leaves, and root of representative healthy ginseng **(A)** and asymptomatic diseased ginsengs **(B–D)**. The dashed green box represents the healthy part, the dashed red box represents the diseased part.

taken, and healthy and diseased ginseng roots were separated by visual observation on whether there was disease spot on the surface or rot part. The leaves were removed from each stem, shortly frozen in the liquid nitrogen and then stored at -80°C. Based on the visually assessed symptoms of ginseng roots, 57 samples were healthy, and 160 samples were diseased (Figure 2).

The Frozen leaves of randomly selected 15 healthy and 15 diseased ginsengs were used for antioxidant enzyme activity measurement. Three leaves of each ginseng (~ 0.2 g) were weighed for these measurements. The superoxide dismutase (SOD) enzyme activity of ginseng leaves was determined by nitro blue tetrazolium photoreduction method, peroxidase (POD) enzyme activity was determined by guaiacol method, and catalase (CAT) enzymes activity was determined by ultraviolet spectrophotometry, as described previously by Li (2000).

2.3 The calculation of spectral vegetation indices

In this study, 8 vegetation indices that related to plant stresses (Ayanlade, 2017; Xue and Su, 2017; Velichkova et al., 2019) were calculated (details see the supplementary, Table S1). These vegetation indices are mainly types of narrowband greenness and leaf pigments. The red and near-infrared region (red edge) between 690-740 nm was used to calculate the narrowband greenness.

2.4 Preprocessing of hyperspectral reflectance

The raw data of hyperspectral reflectance of ginseng leaves were exported by the spectral data processing software View Spec Pro (Analytical Spectral Devices, Colorado, US) from the spectrometer. Machine noises at the beginning and the end of a band were removed according to the fluctuation range of the hyperspectral reflectance curve, and the spectral data with less machine noises were selected for subsequent processing. In this study, 10 preprocessing schemes including first derivative (1D), second derivative (2D), standard normal variate (SNV), multiple scattering correction (MSC), savitzky-golay (SG), 1D-SG, 2D-SG, SNV-SG, and MSC-SG were used, which were completed by the software SIMCA_P+ 13.0 (Umetrics, Umea, Sweden). SNV is

used to correct for light scattering differences. MSC effectively eliminates spectral differences due to different scattering levels. Derivative processing effectively improves sensitivity and resolution. SG preprocessing smoothest high-frequency noise and effectively improves the signal-to-noise ratio. The fluctuation range and smoothness of spectral reflectance were compared according to the image, and the accuracy was compared according to the preliminary modeling results. The best preprocessing method was selected and then further modeling was performed.

2.5 Construction of detection models and evaluation indices

Based on the Python 3.7.0 open-source ML toolkit Scikit-Learn (Pedregosa et al., 2011), models were built using 4 algorithms (RF; ET, extremely randomized trees; ADA, adaptive boosting; GBDT, gradient boosting decision tree) with the hyperspectral dataset of ginseng leaves after optimal preprocessing. The basic unit of RF is a single decision tree, and its main idea is ensemble learning (EL). Every decision tree in a RF is a classifier. For an input sample, the RF algorithm assigns the class with the most votes as the final output by performing an equal vote on the predictions of all decision trees (Breiman, 2001). The ET algorithm is very similar to the RF algorithm. The difference between the two is that the training set of each decision tree in the RF model is obtained by random sampling, while each decision tree in the ET model uses the original training set (Geurts et al., 2006). Therefore, the variance of the ET algorithm is lower than that of RF to some extent, but its bias is relatively high. Furthermore, unlike RF which selects the optimal eigenvalue split point, ET usually randomly selects an eigenvalue split point. ADA is an ensemble algorithm based on the boosting strategy. Its core idea is to use the training set to train multiple weak classifiers. During this process, the weight of the sample and the weight of the classifier are constantly changing. Finally, a strong classifier is formed. The "adaptation" of the ADA algorithm is reflected in the fact that the weight of each sample is determined by the accuracy of the previous prediction (Freund and Schapire, 1997; Bauer and Kohavi, 1999). GBDT is also an EL based on boosting strategy, but it is different from the ADA algorithm. For the GBDT algorithm, it does not have the concept of sample weight, but adopts the concept of "residual error". In detail, the GBDT algorithm is fitted for the negative gradient of the current model, and in this process, the error value of the weak learner is getting smaller and smaller (Friedman, 2001). Since 10 measurements were recorded for one plant, there were a total of 2 170 groups of data for all plants. These data were divided randomly into the training set and test set with the ratio of 3:1. The healthy ginseng was represented by 0, and the diseased ginseng was represented by 1. The training set was used for model establishment and optimization, and the test set was used to test the hyperspectral reflectance data collected from ginseng leaves. Then Grid Search and Learning Curve were combined to adjust the hyperparameters of the model, including n_estimators, max-depth, and subsample, so as to achieve the purpose of optimizing the model (Isa et al., 2019). The optimal hyperparameter combination for model optimization was set as follows: n_estimators = 100, max_depth = 17 for RF model; n_estimators = 50, max_depth = 19 for ET model; n_estimators = 240 for the ADA model; and n_estimators = 270, subsample = 0.7 for the GBDT model.

Finally, the importance of all hyperspectral bands was sorted, and the characteristic bands that played important roles in the early detection of ginseng root diseases were screened out. The performance of each model was compared by calculating different indicators, including accuracy, precision, recall, f1-score, area under the curve (AUC) and Matthew's correlation coefficient (MCC). Among the indicators, precision represents the proportion of truly diseased samples to the samples predicted to be diseased by the model; recall represents the proportion of all diseased samples that the model predicted correctly; and f1-score is the balance coefficient between precision and recall. By drawing receiver operating characteristic curve (ROC) and confusion matrix chart, the performance of the model itself and the overall prediction effect were objectively evaluated. The specific calculation formula are as follows:

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (1)$$

$$Precision = TP/(TP + FP) \quad (2)$$

$$Recall = TP/(TP + FN) \quad (3)$$

$$F1 - score = 2 \times Precision \times Recall/(Precision + Recall) \quad (4)$$

Where TP (true positive) indicates the number of diseased ginsengs correctly predicted, TN (true negative) indicates the number of healthy ginsengs correctly predicted, FP (false positive) indicates the number of healthy ginsengs wrongly predicted, FN (false negative) indicates the number of diseased ginsengs wrongly predicted.

2.6 Model validation

Besides 217 ginseng plants, hyperspectral reflectance of leaves of randomly selected another 12 ginsengs with healthy leaves growing were also collected using HH2 in the same way. The roots of these 12 ginsengs were harvested for the classification of healthy or diseased ginsengs. The selected model was applied to predict these randomly selected 12 ginsengs. Ten groups of hyperspectral reflectance data were collected from each plant. The average ten groups of data of each plant were calculated and was used for the prediction. The predicted results of the model were compared to

the classification after harvest and the prediction accuracy was calculated.

2.7 Statistical analysis

Statistical analysis in this study was conducted using SPSS 20.0 Software (IBM Corp., Armonk, NY, USA). SPAD values, antioxidant enzyme activities, and vegetative indices of ginseng leaves from both healthy and diseased groups were compared for differences using student T-tests.

3 Results

3.1 The chlorophyll content and the activity of antioxidant enzymes

The chlorophyll contents of leaves of healthy ginsengs were slightly higher than those of leaves of diseased ginsengs, but this difference was not significant (Figure 3A). The absence of lesions on leaves inhibits us to recognize root diseases at their early stages. Similarly, the antioxidant enzyme activities of SOD, POD and CAT in leaves of diseased ginsengs were also slightly higher than those in leaves of healthy ginsengs (Figures 3B–D), but not significantly (n=15, P>0.05). Thus, the activities of antioxidant enzymes were in accordance with the macroscopic symptoms of leaves, and cannot be the indicator of root diseases.

3.2 Hyperspectral vegetation indices

VIs related to vegetation vitality, anthocyanin and carotenoid content were calculated with hyperspectral reflectance. The narrowband greenness values, such as red edge normalized difference vegetation index (NDVI), modified red edge simple ratio (MSR), modified NDVI (mNDVI), and vogelmann red edge index1 (VOG1), of diseased ginseng leaves were lower than those of healthy

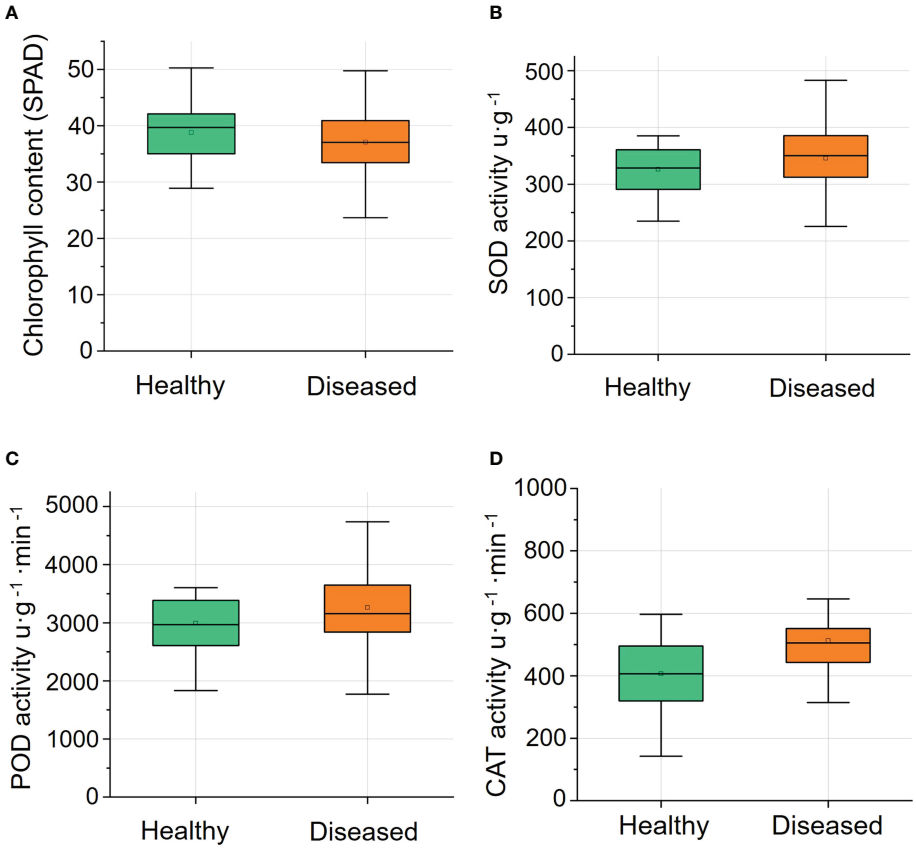


FIGURE 3
The chlorophyll content (n=57) and the activity of antioxidant enzymes (n=15) of leaves of healthy and diseased ginsengs. **(A)** Changes of relative chlorophyll content in healthy and diseased ginseng. **(B)** Changes of SOD activity in healthy and diseased ginseng. **(C)** Changes of POD activity in healthy and diseased ginseng. **(D)** Changes of CAT activity in healthy and diseased ginseng. SPAD, soil plant analysis development; SOD, superoxide dismutase; POD, peroxidase; CAT, catalase.

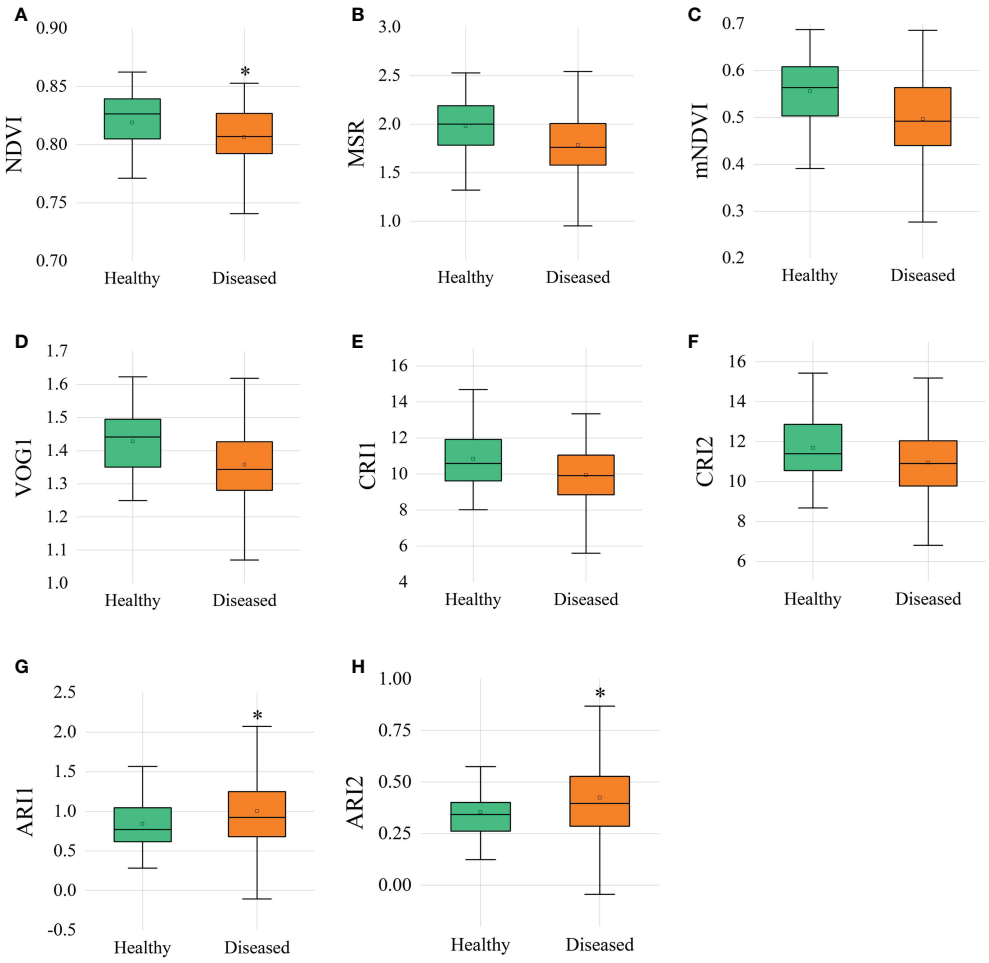


FIGURE 4
Vegetation indices of healthy and diseased ginsengs. n=57, * indicates the significant difference between healthy and diseased (P< 0.05). **(A)** NDVI, red edge normalized difference vegetation index; **(B)** MSR, modified red edge simple ratio; **(C)** mNDVI, modified NDVI; **(D)** VOG, vogelmann red edge index; **(E, F)** CRI, carotenoid reflectance index; **(G, H)** ARI, anthocyanin reflectance index.

ginsengs (Figure 4). Especially the value of NDVI, which was significantly lower (P<0.05) in the diseased ginsengs than in the healthy ginsengs, indicating that at the position of the red edge (690-740 nm), ginsengs with diseased roots could be identified by the NDVI value. In addition, the leaf pigments of diseased ginseng leaves, represented by anthocyanin reflectance index 1 (ARI1) and ARI2, showed significant higher values (P<0.05) than those of healthy ginseng leaves (Figure 4), indicating that the content of anthocyanins in diseased ginseng leaves had been increased, and senescence symptoms would be expected to appear in the next stage of the diseases. However, the VIs related to carotenoids, such as carotenoid reflectance index 1 (CRI1) and CRI2 values of diseased ginsengs, were slightly lower, but not significantly changed, than those of the healthy ginsengs (Figure 4).

3.3 Hyperspectral reflectance

3.3.1 Pre-processing of hyperspectral reflectance data

Hyperspectral bands and a large amount of reflectance data were obtained from ginseng leaves. We selected hyperspectral data with less noises in the region of 460-950 nm by removing the noises at both ends of hyperspectral bands (Figure 5). Among the 10 pre-processing methods (single methods of SNV, MSC, derivative and SG and their different combinations), the SG retained the trend of the original spectral curve, and the reflectivity range was concentrated in 0-0.5 with a clear curve outline (Figure S1).

The accuracy rates of RF, ADA, GBDT, and ET under the raw dataset conditions were 94.81%, 80.08%, 84.65%, and

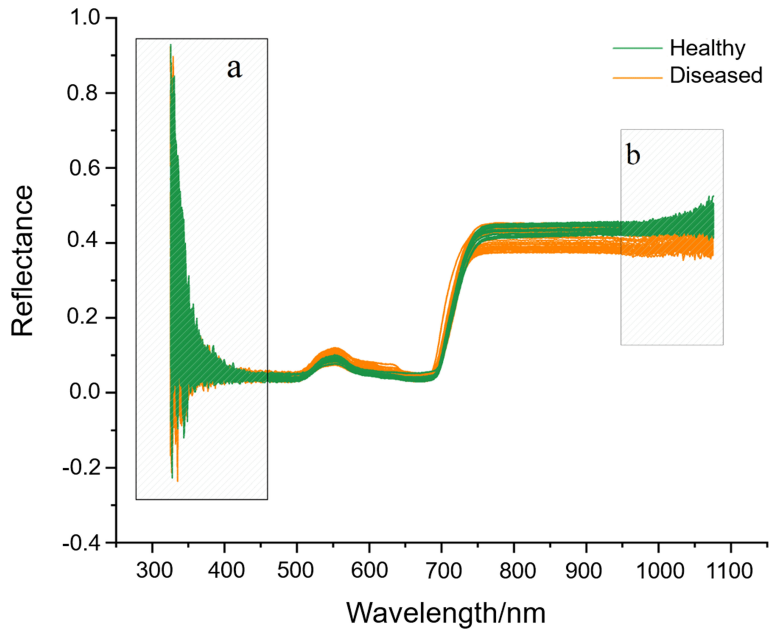


FIGURE 5
The original hyperspectral reflectance of ginseng leaves **(A, B)** are the band ranges with machine noises and are visually selected to be removed.

97.13%, respectively (Table 1). Among the 10 pre-processing methods, after SG pre-processing, the models of RF, GBDT, and ET achieved the highest scores, which were 95.76%, 87.29%, and 97.97%, respectively. After SG pre-processing, the ADA model's accuracy was 79.37%, which was a bit lower than the raw data and the data pre-processed after SNV-SG. Thus, based on the accuracy of the models and the changes of reflectivity curve, the method of SG for data pre-processing was taken as the best pre-processing method in this study.

3.3.2 Model optimization and evaluation

Model optimization improves the recognition ability of the models. The discriminable ability of the 4 models before and after model optimization were listed in Table S2. After parameter adjustment, the accuracy of the ADA and GBDT were improved. The accuracy of Boosting's ADA model was 88%, which was the lowest, indicating the classification effect of this model was slightly worse than the others. Compared with ADA, the accuracy of GBDT model was 4 percent higher, and the bagging strategy-based RF algorithm was 8 percent higher. ET model had the best performance with accuracy of 98%, precision of 98%, recall of 100% and F1-Score of 99%, suggesting it to be the appropriate model for classification. According to the ROC curves (Figure 6A), the AUC of the model of RF and ET, which were based on the bagging strategy, were relatively high, reaching 98.8% and 99.8%, respectively; and the AUC of the models of ADA and GBDT, which were based on the boosting strategy, were relatively low, reaching 84.1% and 97.0%, respectively. It was worth noting that the bagging strategy had an advantage over the boosting strategy on this dataset. According to the confusion matrices (Figure 6B), the overall prediction effect of all models for class 1 was better than class 0,

TABLE 1 Accuracy rates of 10 preprocessing methods based on the 4 algorithms.

Accuracy (%)	Raw	1D	2D	SNV	MSC	SG	1D-SG	2D-SG	SNV-SG	MSC-SG
RF	94.81	86.56	78.64	87.66	86.92	95.76	91.97	92.45	87.58	79.23
ADA	80.08	77.35	72.38	78.45	77.72	79.37	79.19	78.82	81.49	72.46
GBDT	84.65	83.24	78.27	82.69	82.87	87.29	86.56	86.74	86.00	75.62
ET	97.13	89.32	78.27	91.53	91.16	97.97	96.69	96.13	92.10	80.14

RF, Random forest; ET, Extremely randomized trees; ADA, Adaptive boosting; GBDT, Gradient boosting decision tree; RAW, Raw dataset; 1D, first derivative; 2D, Second derivative; SNV, Standard normal variate; MSC, Multiple scattering correction; SG, Savitzky golay; 1D-SG, first derivative- savitzky golay; 2D-SG, Second derivative – savitzky golay; SNV-SG, Standard normal variate – savitzky golay; MSC-SG, Multiple scattering correction- savitzky golay.

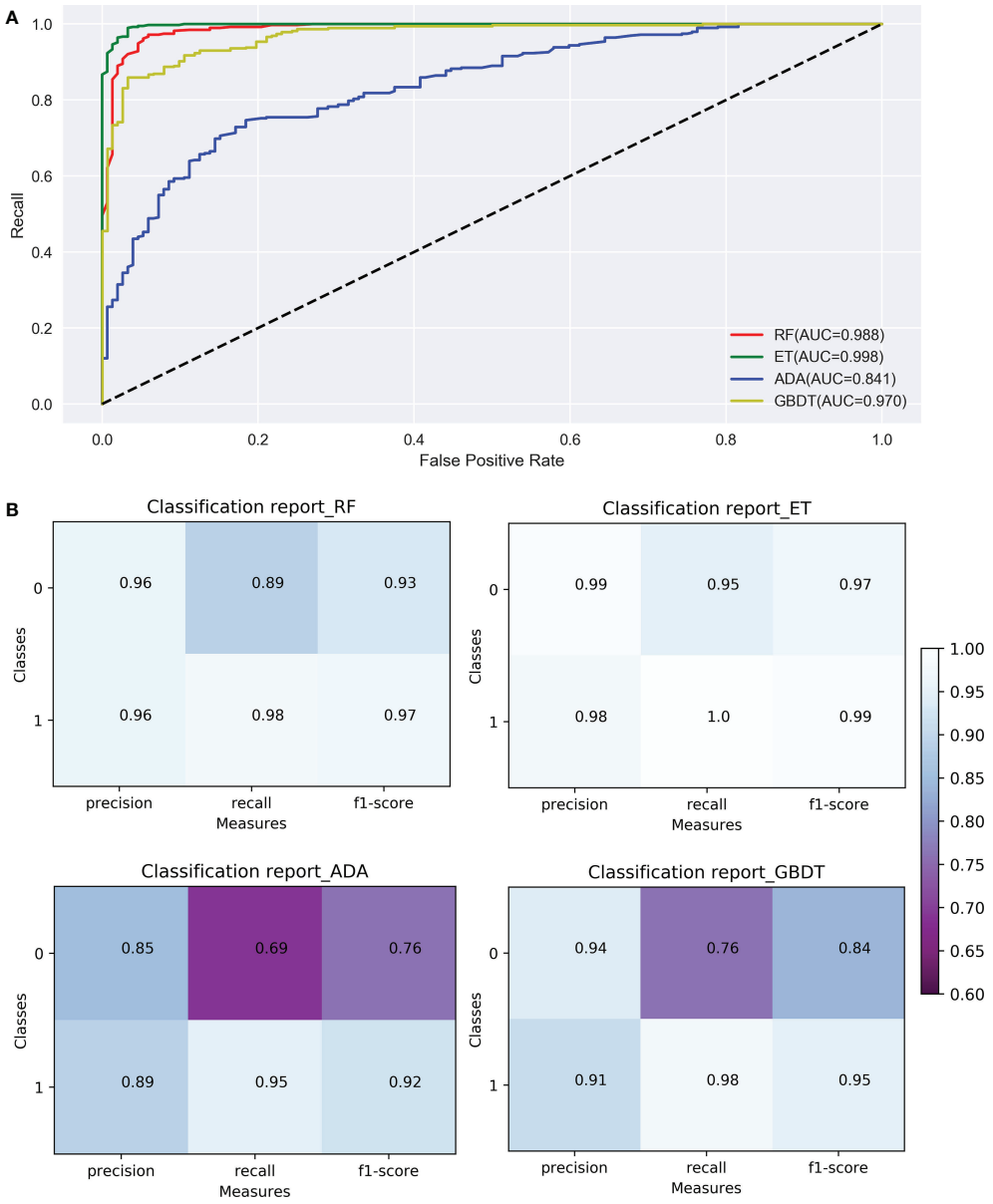


FIGURE 6
Comparison of ROC curves **(A)** and confusion matrix **(B)** of the 4 models (RF, ET, ADA, GBDT). RF, Random Forest; ET, Extremely randomized trees; ADA, Adaptive Boosting; GBDT, Gradient Boosting Decision Tree; Classes: 0 indicates healthy ginseng, 1 indicates diseased ginseng.

indicating that it had a better recognition effect on ginseng root disease. The f1-score of all models in class 1 were greater than 0.9 (Figure 6B), meaning that all models achieved a balance between precision and recall. Among the four models, the ET model has the highest f1-score of 0.99, precision of 0.98 and recall of 1.0 (Figure 6B), which was taken to be the optimal model for evaluation the early detection of ginseng root diseases. Similar to the results of AUC, the models of the bagging strategy (RF and ET) were better than the models of the boosting strategy (GBDT and ADA) on this dataset.

3.3.3 The determination of the range of hyperspectral bands of importance

The hyperspectral bands of ginseng leaves measured in this experiment ranged from 325 to 1 075 nm, with a total of 750 bands. After pre-processing, 490 bands were selected as 490 variables. After processing these 490 variables by RF, ET, ADA, and GBDT algorithms, it was concluded that the most important bands for each model were 691 nm, 726 nm, 714 nm, and 716 nm, respectively (Figure 7). The most important bands were all ranged between 690-726 nm (Figure 7). Although the high

10.3389/fpls.2022.1031030

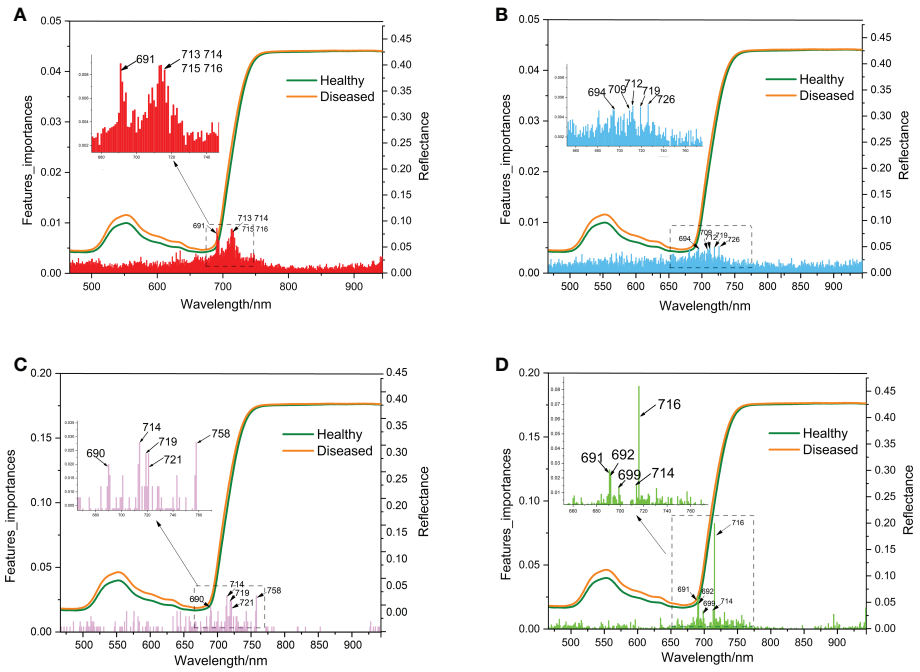


FIGURE 7
Schematic diagram of wavelength importance analysis of the 4 models (**A**: RF, **B**: ET, **C**: ADA, **D**: GBDT) same as Figure 6.

randomness of the RF and ET models resulted in low band importance scores (less than 0.01), 690-726 nm was also identified within the top 5 bands of importance for these two models (Figure 7). As a result, in the hyperspectral reflectance of ginseng leaves, the differences of bands close to 690-726 nm were most likely to indicate whether a ginseng root was diseased or not, and this range of bands can be considered as the hyperspectral characteristic bands for ginseng root disease detection.

3.4 Model validation

According to the evaluation of various indicators of RF, ET, ADA and GBDT models, the ET model was selected as the early detection model for ginseng root diseases. The hyperspectral reflectance of leaves (with no symptoms) of another 12 plants growing in the nearby field were collected and were used for model validation. Among the 12 plants, 9 were detected to be diseased ones and the other 3 were considered to be healthy

TABLE 2 Root diseases prediction of 12 ginsengs in another plot by the ET model.

No. of Ginsengs	Prediction by the model	Detection after sampling	Correctness of the model
1	0	Di	No
2	1	Di	Yes
3	1	Di	Yes
4	1	Di	Yes
5	1	Di	Yes
6	1	Di	Yes
7	0	Di	No
8	1	Di	Yes
9	1	Di	Yes
10	1	Di	Yes
11	1	Di	Yes
12	0	He	Yes

Model prediction: 0 refers to the healthy root, whereas 1 refers to the diseased root. After sampling, whether the root of each ginseng was healthy (He) or diseased (Di), was identified by vision.

according to the models (Table 2). After harvest, these 12 ginsengs were diagnosed visually (Figure S2), 11 were found to be diseased and 1 was healthy (Table 2). The model correctly identified the 9 diseased ginsengs and the 1 healthy ginseng. Only two plants (No. 1 and 7, Table 2), which were actually diseased ones, were mistakenly predicted to be the healthy ones. Thus, the accuracy of this model for these 12 ginsengs is 83.3%.

4 Discussion

The development of root disease is a gradual process. It is generally believed that symptoms on leaves appear much later until the diseases of the root are severe to a certain extent (Zhou et al., 2017). Surprisingly, our research found that when the root is seriously rotted, there are still ginsengs with leaves showing no symptoms (Figure 2D). Along with the visual observation, there were no significant changes in the physiological traits, like SPAD and antioxidant enzyme activities (Figure 3). The insignificantly changed SPAD indicated that the chlorophyll contents of leaves were not affected and even the photosynthesis system may still be well functioning when roots were infected. Even VIs including MSR, mNDVI1, VOG1, CRI1 and CRI2 were not significantly affected in these leaves, confirming the changes of leaves caused by root diseases were very slight. It is not clear how leaves could keep healthy when roots are already severely rotted, but it reminds us that the macro symptoms and even physiological traits of leaves may not reflect the status of roots and a more accurate detection method such as the one based on the spectroscopic information is necessary.

The application of hyperspectral remote sensing technology combined with ML is prospectively used in detecting diseases of an increasing number of crops (Lowe et al., 2017; Yang et al., 2020; Tian et al., 2021). So far, this technology has not been widely used in disease detection of medicinal plants. Most disease detection and identification of crops or other plants are performed on images of plant tissues where the diseases occur with visible symptoms (Singh et al., 2018; Singh et al., 2021), such as leaf blast and false smut infection of rice (An et al., 2021; Tian et al., 2021) and apple fire blight disease (Jarolmasjed et al., 2019). However, the medicinal parts of most medicinal plants, like ginsengs, are roots, whose hyperspectral reflectance data cannot be collected directly and non-constructively. Calamita et al. (2021) found significant differences in the near-infrared spectral region of leaves between healthy and root-rot grape plants by the naive bayes algorithm, which showed 90% accuracy in the identification of healthy and diseased plants, indicating the possibility for the diagnosis of root rot in plants by applying hyperspectral reflectance from leaves. However, such detection method is rarely mentioned by previous researchers and has not been reported in medicinal plants. In this study, we established a method to detect ginseng root diseases by collecting and analysing the hyperspectral reflectance data of leaves. Since the relationship between leaves and diseased root is still unclear, this indirect detection highly demands on the precision of data collection and analysis. Thus, the recognition of the region of characteristic bands and a proper method of ML processing are very important.

We found visible (460-760 nm) and near-infrared (760-950 nm) spectral reflectance of leaves played an important role in monitoring ginseng growth (Figure 5). Based on the data collected from leaves from this region and the visual detection of ginseng roots after harvest, ML algorithms were used to construct the model of root diseases detection. The combination of non-destructively acquired hyperspectral reflectance data and ML algorithms can recognize the tiny changes of hyperspectral reflectance of the asymptomatic patients, that greatly improves the classification accuracy of the model (Sankaran et al., 2012; Abdulridha et al., 2022). For example, the logistic regression-based ML algorithms by Appeltans et al. (2021) obtained the accuracy of automatically labelled *Phytophthora infestans* is 98.80%, wheat *Puccinia striiformis* and *Puccinia triticina* are 97.69% and 96.66%, respectively. The four ML models of RF, ET, ADA, and GBDT used in this study all belonged to EL have achieved high accuracy in the detection of ginseng root diseases (> 85%). EL is a commonly used ML algorithm in processing hyperspectral datasets (Wei et al., 2020; Ekramirad et al., 2022). Its advantage is to organize several simple algorithms to jointly determine the final performance. Among the models, RF and ET are based on bagging strategy and mainly optimize the robustness (variance) of the model (Breiman, 2001; Svetnik et al., 2003). For example, RF hyperspectral model can better predict heavy metal distribution in agricultural soil (Tan et al., 2020). Whereas ADA and GBDT are based on boosting strategy and mainly optimize the precision (bias) of the model (Freund and Schapire, 1997). For example, the GBDT model can validly classify apple bruising times (Pan et al., 2019). In this study, after parameter adjustment, the models of the bagging strategy (RF and ET) were more suitable for this dataset. The accuracy of ET model was highest of 98% and the AUC of ET model is as high as 99.8% (Figure 6). This is consistent with the good generalization ability of the ET algorithm, that successfully classified seven types of Spanish honeys with single botanical origins (Mateo et al., 2021).

Besides model selection, the pre-processing method is also strongly affected the model accuracy. The partial least squares discriminant analysis achieves the best results in SNV-processed *Paris yunnanensis* data of different origins (Pei et al., 2018). For the ET model in this study, the accuracy of the SG pre-processing method is 19.52 percentage higher than that of the 2D pre-processing method (Table 1). Based on the contribution rate of the models, the most important bands collected from ginseng leaves were all concentrated in the range of 690-726 nm (Figure 7). Combined with the analysis of hyperspectral VIs (Figure 4), the significant differences of NDVI were also

associated with this range. The reflectance at the red edge position is suggested to be used to evaluate the structural changes and physiological degradation of leaves (Vescovo et al., 2012; Liu et al., 2014). Thus, the changes in the position of red edges in ginseng leaves may be the results of unobservable changes in leaf structure caused by lesions of ginseng root. Therefore, the range of characteristic bands for early detection of ginseng root diseases was further narrowed.

This study explored the detection method of root diseases based on the hyperspectral reflectance of asymptomatic ginseng leaves. The method was built based on the combination of non-destructively acquired hyperspectral data from healthy leaves and the ML classification algorithms. After pre-processing and optimizing the data set, classification models combined with the indicators were evaluated. Finally, the method for early detection of ginseng root disease was built and validated. The pipeline of the construction of the method was shown in Figure 8. Besides, ginsengs from another plot with healthy leaves were also collected and were used to validate the model. Though with a limited number of samples, the model achieves a correct rate of 83.3%, showing the effectiveness of the method. Upon the high detection rate of the method, the method can still be improved in many aspects. The hyperspectral reflectance can be affected by many factors, such as the ages of ginseng, the development stages of ginsengs, the types or causes of diseases, the severity of diseases as well as the changes of environments. Thus, based on the effectiveness of the simple model we built in this study, it is promising to develop an integrated algorithm based on data collected during the precise identification phase of root disease development. In addition to the two groups of healthy and diseased, more variables such as the development stages of ginsengs, the environmental factors, the type of the diseases and the severity of the diseases can be brought into the model. To be more precisely, hyperspectral data can be collected in the manual interventions conditions, such as different stages of ginsengs being pathogenic inoculated or ginsengs growing under certain abiotic stresses, that lead to a certain type of ginseng root diseases. Thus, a lot more data are expected to be

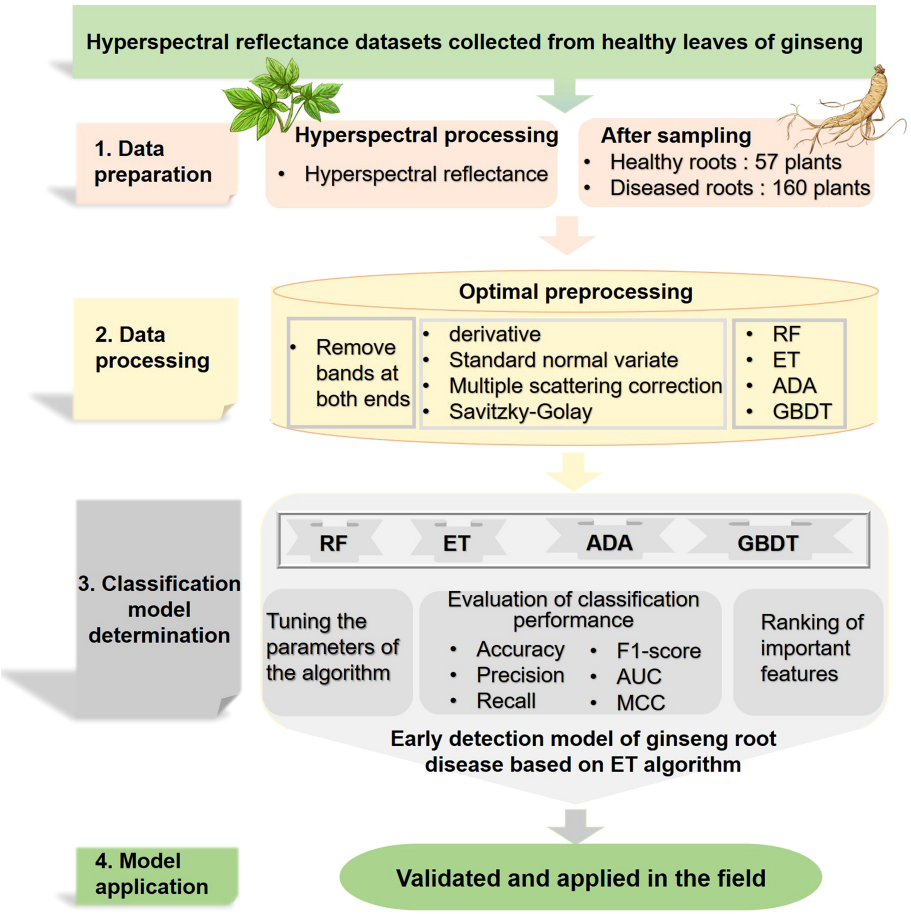


FIGURE 8
Schematic diagram of the classification method based on the ET algorithm for spectral detection of ginseng root diseases. same as Figure 6; AUC, area under the curve; MCC, Matthew's correlation coefficient.

collected, and a hyperspectral reflectance database of ginseng root diseases database can be built. Since the training of ML models requires large amount of data (Tsaftaris et al., 2016), comprehensive algorithms are supposed to be developed and improved based on the database. In addition, the development of the detection method could be combined with pathological studies, which would provide basic knowledge of the mechanism how hyperspectral reflectance are related to the diseased roots, aiding the construction of ML models. To summarize, based on the detection method of ginseng root diseases constructed by the hyperspectral reflectance of asymptomatic ginseng leaves that has been proposed in this study, a comprehensive hyperspectral reflectance database is expected to be built in the future together with the development of ML models, in order to accurately identify the type, onset time, severity of underground diseases of ginsengs and to perform timely treatments to reduce production loss. Early detection of root diseases is essential as timely excavation of diseased ginseng in the early stages of the disease can prevent further infection of other plants. This will help to provide new methods for disease detection in other root plants and also provide ideas for the development of more portable and simple field detection equipment. Equipping hyperspectral remote sensing equipment with intelligent robots for self-service detection will also be a future goal for the development of smart agriculture in the field. By establishing abiotic stress tests, combined with hyperspectral remote sensing technology to further explore the optimal cultivation conditions for ginseng root growth, the environmental conditions for ginseng root disease development will be blocked at source and root disease prevention will be achieved. In the field cultivation of ginseng, the diseased ginseng can be harvested according to the predicted results based on our established model for early detection of root diseases, further completing the identification of ginseng root disease pathogens, exploring the pathogenic mechanisms of fungi and bacteria, screening for efficient and non-polluting antibacterial substances, and finally achieving the prevention and control of the disease.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/yang1rq/ginseng/tree/master.

Author contributions

XL, DaY, and GZ conceived, designed, processed, analysed, and interpreted the experiments. GZ, DoY, and JW acquired the data. GZ, YP and RY analysed the data and prepared the manuscript. GZ, LX, ZF, YW and DG edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The research was funded by National Key Research and Development Program of China (No. 2019YFC1710601), Scientific and technological innovation project of China Academy of Chinese Medical Sciences (CI2021A04505), National Natural Science Foundation of China (No. 81891013) and the Fundamental Research Funds for the Central public welfare research institutes (ZZ15-YQ-031, ZXKT20133).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1031030/full#supplementary-material

References

Żelazny, W. R., Chrpová, J., and Hamouz, P. (2021). Fusarium head blight detection from spectral measurements in a field phenotyping setting — a pre-registered study. *Biosyst. Eng.* 211, 97–113. doi: 10.1016/j.biosystemseng.2021.08.019

Abdulridha, J., Ampatzidis, Y., Qureshi, J., and Roberts, P. (2022). Identification and classification of downy mildew severity stages in watermelon utilizing aerial and ground remote sensing and machine learning. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.791018

Adavi, Z., Tadayon, M. R., Razmjoo, J., and Ghaffari, H. (2020). Antioxidant enzyme responses in potato (*solanum tuberosum*) cultivars colonized with arbuscular mycorrhizas. *Potato Res.* 63, 291–301. doi: 10.1007/s11540-019-09440-1

An, G., Xing, M., He, B., Kang, H., Shang, J., Liao, C., et al. (2021). Extraction of areas of rice false smut infection using uav hyperspectral data. *Remote Sens. (Basel Switzerland)* 13, 3185. doi: 10.3390/rs13163185

Appeltans, S., Apolo-Apolo, O. E., Rodríguez-Vázquez, J. N., Pérez-Ruiz, M., Pieters, J., and Mouazen, A. (2021). The automation of hyperspectral training library construction: a case study for wheat and potato crops. *Remote Sens.* 13, 4735. doi: 10.3390/rs13234735

Arya, S., Sandhu, K. S., Singh, J., and Kumar, S. (2022). Deep learning: as the new frontier in high-throughput plant phenotyping. *Euphytica* 47, 218. doi: 10.1007/s10681-022-02992-3

Ayanlade, A. (2017). Remote sensing vegetation dynamics analytical methods: a review of vegetation indices techniques. *Geoinformatica Polonica* 16, 7–17. doi: 10.4467/21995923GP.17.001.7188

Bauer, E., and Kohavi, R. (1999). An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Mach. Learn.* 36, 105–139. doi: 10.1023/A:1007515423169

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Calamita, F., Imran, H. A., Vescovo, L., Mekhalfi, M. L., and La Porta, N. (2021). Early identification of root rot disease by using hyperspectral reflectance: The case of pathosystem grapevine/*armillaria*. *Remote Sens.* 13, 2436. doi: 10.3390/rs13132436

Chen, S., Li, Z., Zhang, S., Zhou, Y., Xiao, X., Cui, P., et al. (2022). Emerging biotechnology applications in natural product and synthetic pharmaceutical analyses. *Acta Pharm Sin B* 12, 4075–4097. doi: 10.1016/j.apsb.2022.08.025

Deng, L., Feng, G., Gao, Y., Shen, Y., Li, H., Gu, Y., et al. (2020). Phytochemical constituents and antioxidant enzyme activity profiles of different barley (*hordeum vulgare* l.) cultivars at different developmental stages. *Agronomy* 10, 37. doi: 10.3390/agronomy10010037

Ekramirad, N., Khaled, A. Y., Doyle, L. E., Loeb, J. R., Donohue, K. D., Villanueva, R., et al. (2022). Nondestructive detection of codling moth infestation in apples using pixel-based nir hyperspectral imaging with machine learning and feature selection. *Foods* 11, 8. doi: 10.3390/foods11010008

Fang, X., Wang, H., Zhao, L., Wang, M., and Sun, M. (2022). Diversity and structure of the rhizosphere microbial communities of wild and cultivated ginseng. *BMC Microbiol.* 22, 2. doi: 10.1186/s12866-021-02421-w

Farh, M. E., Kim, Y., Kim, Y., and Yang, D. (2018). *Cylindrocarpon destructans/ilyonectria* radicicola-species complex: causative agent of ginseng root-rot disease and rusty symptoms. *J. Ginseng Res.* 42, 9–15. doi: 10.1016/j.jgr.2017.01.004

Feng, Z., Wang, L., Yang, Z., Zhang, Y., Li, X., Song, L., et al. (2022). Hyperspectral monitoring of powdery mildew disease severity in wheat based on machine learning. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.828454

Freund, Y., and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* 55, 119–139. doi: 10.1006/jcss.1997.1504

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine *Annal Stat.* 29, 1189–1232. doi: 10.1214/AOS/1013203451

Fujimoto, T., Abe, H., Mizukubo, T., and Seo, S. (2021). Phytol, a constituent of chlorophyll, induces root-knot nematode resistance inarabidopsis *via* the ethylene signaling pathway. *Mol. Plant-Microbe Interactions®* 34, 279–285. doi: 10.1094/MPMI-07-20-0186-R

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1

Ghosh, D., Chakraborty, S., Kodamana, H., and Chakraborty, S. (2022). Application of machine learning in understanding plant virus pathogenesis: Trends and perspectives on emergence, diagnosis, host-virus interplay and management. *Virol. J.* 19, 42. doi: 10.1186/s12985-022-01767-5

Guan, Y. M., Lu, B. H., Wang, Y., Gao, J., and Wu, L. J. (2014). First report of root rot caused by *fusarium* redolens on ginseng (*panax ginseng*) in jilin province of China. *Plant Dis.* 98, 844. doi: 10.1094/PDIS-03-14-0328-PDN

Harrach, B. D., Baltruschat, H., Barna, B., Fodor, J., and Kogel, K. (2013). The mutualistic fungus *piriformospora indica* protects barley roots from a loss of antioxidant capacity caused by the necrotrophic pathogenfusarium culmorum. *Mol. Plant-Microbe Interactions®* 26, 599–605. doi: 10.1094/MPMI-09-12-0216-R

Isa, S. M., Suwandi, R., and Pricilia, Y. (2019). Optimizing the hyperparameter of feature extraction and machine learning classification algorithms. *Int. J. Advanced Comput. Sci. Appl.* 10, 69–76. doi: 10.14569/IJACSA.2019.0100309

Jarolmasjed, S., Sankaran, S., Marzougui, A., Kostick, S., Si, Y., Vargas, J., et al. (2019). High-throughput phenotyping of fire blight disease symptoms using sensing techniques in apple. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00576

Jia, Y., Huang, J., Qi, L., Zhang, X., Liu, J., Guan, H., et al. (2021). *Bacillus subt*ilis strain bs06 protects soybean roots from *fusarium oxysporum* infection. *FEMS Microbiol. Lett.* 368, 1–7. doi: 10.1093/femsle/fnab102

Koh, J. C. O., Banerjee, B. P., Spangenberg, G., and Kant, S. (2022). Automated hyperspectral vegetation index derivation using a hyperparameter optimisation framework for high-throughput plant phenotyping. *New Phytol.* 233, 2659–2670. doi: 10.1111/nph.17947

Li, H. (2000). *Principles and techniques of plant physiological biochemical experiment* (Beijing, China: Higher Education Press).

Liu, L., Huang, W., PU, R., and Wang, J. (2014). Detection of internal leaf structure deterioration using a new spectral ratio index in the near-infrared shoulder region. *J. Integr. Agric.* 13, 760–769. doi: 10.1016/S2095-3119(13)60385-8

Lowe, A., Harrison, N., and French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* 13, 80. doi: 10.1186/s13007-017-0233-z

Lu, X. H., Zhang, X. M., Jiao, X. L., Hao, J. J., Zhang, X. S., Luo, Y., et al. (2020). Taxonomy of fungal complex causing red-skin root of *panax ginseng* in china. *J. Ginseng Res.* 44, 506–518. doi: 10.1016/j.jgr.2019.01.006

Mahlein, A. K., Kuska, M. T., Behmann, J., Polder, G., Walter, A., and Rascher, U. (2018). Hyperspectral sensors and imaging technologies in phytopathology: State of the art. *Annu. Rev. Phytopathol.* 56, 535–558. doi: 10.1146/annurev-phyto-080417-050100

Mahlein, A., Kuska, M. T., Thomas, S., Wahabzada, M., Behmann, J., Rascher, U., et al. (2019). Quantitative and qualitative phenotyping of disease resistance of crops by hyperspectral sensors: Seamless interlocking of phytopathology, sensors, and machine learning is needed! *Curr. Opin. Plant Biol.* 50, 156–162. doi: 10.1016/j.pbi.2019.06.007

Mateo, F., Tarazona, A., and Mateo, E. M. (2021). Comparative study of several machine learning algorithms for classification of unifloral honeys. *Foods* 10, 1543. doi: 10.3390/foods10071543

Monteoliva, M. I., Guzzo, M. C., and Posada, G. A. (2021). Breeding for drought tolerance by monitoring chlorophyll content. *Gene Technol.* 10, 165. doi: 10.35248/2329-6682.21.10.165

Panda, D., and Sarkar, R. K. (2013). Natural leaf senescence: Probed by chlorophyll fluorescence, co_2 photosynthetic rate and antioxidant enzyme activities during grain filling in different rice cultivars. *Physiol. Mol. Biol. Plants* 19, 43–51. doi: 10.1007/s12298-012-0142-6

Pan, X., Sun, L., Li, Y., Che, W., Ji, Y., Li, J., et al. (2019). Non-destructive classification of apple bruising time based on visible and near-infrared hyperspectral imaging. *J. Sci. Food Agric.* 99, 1709–1718. doi: 10.1002/jsfa.9360

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.48550/arXiv.1201.0490

Pei, Y., Zhang, Q., Zuo, Z., and Wang, Y. (2018). Comparison and identification for rhizomes and leaves of *paris yunnanensis* based on fourier transform mid-infrared spectroscopy combined with chemometrics. *Molecules* 23, 3343. doi: 10.3390/molecules23123343

Qalavand, F., Esfahani, M. N., Vatandoost, J., and Azarm, D. A. (2022). Enzyme activity and population genetic structure analysis in wheat associated with resistance to bipolaris sorokiniana-common root rot diseases. *Phytochemistry* 200, 113208. doi: 10.1016/j.phytochem.2022.113208

Ratan, Z. A., Haidere, M. F., Hong, Y. H., Park, S. H., Lee, J., Lee, J., et al. (2021). Pharmacological potential of ginseng and its major component ginsenosides. *J. Ginseng Res.* 45, 199–210. doi: 10.1016/j.jgr.2020.02.004

Sankaran, S., Ehsani, R., Inch, S. A., and Ploetz, R. C. (2012). Evaluation of visible-near infrared reflectance spectra of avocado leaves as a non-destructive sensing tool for detection of laurel wilt. *Plant Dis.* 96, 1683–1689. doi: 10.1094/PDIS-01-12-0030-RE

Shang, Y., Chen, Z., Si, Y., and Teng, B. (1996). Research on the causes of red-skin disease of *panax ginseng*. *Chin. Med. Mater* 04, 163–165. doi: 10.13863/j.issn1001-4454.1996.04.001

Shen, L., Li, X., Xu, J., Dong, L., Zhang, N., Naoli, F., et al. (2017). Research on technology regulation of non-polluted cultivation and strategies of panax ginseng cultivation in farmland. *China J. Chin. Materia Med.* 42, 3267–3274. doi: 10.19540/j.cnki.cjcmm.2017.0130

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., Mueller, D., Sandhu, K., et al. (2021). Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* 26, 53–69. doi: 10.1016/j.tplants.2020.07.010

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., and Feuston, B.. (2003). Random forest: A classification and regression tool for compound classification and qsar modeling. *J. Chem. Inf. Comput. Sci.* 43, 1947–1958. doi: 10.1021/ci034160g

Tan, K., Wang, H., Chen, L., Du, Q., Du, P., and Pan, C. (2020). Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest. *J. Hazard Mater* 382, 120987. doi: 10.1016/j.jhazmat.2019.120987

Tian, L., Xue, B., Wang, Z., Li, D., Yao, X., Cao, Q., et al. (2021). Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection. *Remote Sens. Environ.* 257, 112350. doi: 10.1016/j.rse.2021.112350

Tsaftaris, S. A., Minervini, M., and Scharr, H. (2016). Machine learning for plant phenotyping needs image processing. *Trends Plant Sci.* 21, 989–991. doi: 10.1016/j.tplants.2016.10.002

Velichkova, K., Krezhova, D., Mishonov, T. M., and Varonov, A. M. (2019). "Extraction of the red edge position from hyperspectral reflectance data for plant stress monitoring," in *AIP Conference Proceedings*, Vol. 2075, 130018. doi: 10.1063/1.5091303

Vescovo, L., Wohlfahrt, G., Balzarolo, M., Pilloni, S., Sottocornola, M., Rodeghiero, M., et al. (2012). New spectral vegetation indices based on the near-infrared shoulder wavelengths for remote detection of grassland phytomass. *Int. J. Remote Sens.* 33, 2178–2195. doi: 10.1080/01431161.2011.607195

Wang, Q., Jin, R., Yun, L., Wang, J., Jiang, X., Ma, D., et al. (2014). Occurrence investigation and pathogen study on ginseng rusty root rot. *J. Liaoning Forestry Sci. & Technol.* 3, 15–17. doi: 1001-1714(2014)03-0015-03

Wei, Y., Li, X., Pan, X., and Li, L. (2020). Nondestructive classification of soybean seed varieties by hyperspectral imaging and ensemble machine learning algorithms. *Sensors* 20, 6980. doi: 10.3390/s20236980

Xiao, C., Yang, L., Zhang, L., Liu, C., and Han, M. (2016). Effects of cultivation ages and modes on microbial diversity in the rhizosphere soil of *panax ginseng*. *J. Ginseng Res.* 40, 28–37. doi: 10.1016/j.jgr.2015.04.004

Xu, X., Cheng, X., Lin, Q., Li, S., Jia, Z., Han, T., et al. (2016). Identification of mountain-cultivated ginseng and cultivated ginseng using uplc/oa-tof mse with a multivariate statistical sample-profiling strategy. *J. Ginseng Res.* 40, 344–350. doi: 10.1016/j.jgr.2015.11.001

Xue, J., and Su, B. (2017). Significant remote sensing vegetation indices: a review of developments and applications. *J. Sensors* 2017, 1–17. doi: 10.1155/2017/1353691

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008

Zhang, H., Chen, Y., Liu, X., Huang, Y., Zhan, B., and Luo, W. (2021). Identification of common skin defects and classification of early decayed citrus using hyperspectral imaging technique. *Food Analytical Methods* 14, 1176–1193. doi: 10.1007/s12161-020-01960-8

Zhang, D., Wang, Q., Lin, F., Yin, X., Gu, C., and Qiao, H. (2020). Development and evaluation of a new spectral disease index to detect wheat fusarium head blight using hyperspectral imaging. *Sensors* 20, 2260. doi: 10.3390/s20082260

Zhou, Y., Yang, Z., Gao, L., Liu, W., Liu, R., Zhao, J., et al. (2017). Changes in element accumulation, phenolic metabolism, and antioxidative enzyme activities in the red-skin roots of *panax ginseng*. *J. Ginseng Res.* 41, 307–315. doi: 10.1016/j.jgr.2016.06.001

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 November 2022
DOI 10.3389/fpls.2022.1031849

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Ebrahiema Arendse,
Agricultural Research Council of South Africa (ARC-SA), South Africa
Xi Qiao,
Agricultural Genomics Institute at Shenzhen (CAAS), China

*CORRESPONDENCE
Qun Sun
sqcau@126.com
Xuehui Dong
xuehuidong@cau.edu.cn

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 30 August 2022
ACCEPTED 08 November 2022
PUBLISHED 29 November 2022

CITATION
Xu Y, Wu W, Chen Y, Zhang T, Tu K, Hao Y, Cao H, Dong X and Sun Q (2022) Hyperspectral imaging with machine learning for non-destructive classification of *Astragalus membranaceus* var. *mongholicus*, *Astragalus membranaceus*, and similar seeds.
Front. Plant Sci. 13:1031849.
doi: 10.3389/fpls.2022.1031849

COPYRIGHT
© 2022 Xu, Wu, Chen, Zhang, Tu, Hao, Cao, Dong and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Hyperspectral imaging with machine learning for non-destructive classification of *Astragalus membranaceus* var. *mongholicus*, *Astragalus membranaceus*, and similar seeds

Yanan Xu[1], Weifeng Wu[1], Yi Chen[1], Tingting Zhang[2], Keling Tu[1], Yun Hao[1], Hailu Cao[3], Xuehui Dong[1*] and Qun Sun[1*]

[1]College of Agronomy and Biotechnology, Department of Plant Genetics & Breeding and Seed Science/Chinese Medicinal Herbs Research Center, China Agricultural University/The Innovation Center (Beijing) of Crop Seeds whole-process Technology Research, Ministry of Agriculture and Rural Affairs/Beijing Key Laboratory of Crop Genetic Improvement, Beijing, China, [2]Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China, [3]Hengde Materia Medica (Beijing) Agricultural Technology Co., Ltd., Beijing, China

The roots of *Astragalus membranaceus* var. *mongholicus* (AMM) and A. membranaceus (AM) are widely used in traditional Chinese medicine. Although AMM has higher yields and accounts for a larger market share, its cultivation is fraught with challenges, including mixed germplasm resources and widespread adulteration of commercial seeds. Current methods for distinguishing Astragalus seeds from similar (SM) seeds are time-consuming, laborious, and destructive. To establish a non-destructive method, AMM, AM, and SM seeds were collected from various production areas. Machine vision and hyperspectral imaging (HSI) were used to collect morphological data and spectral data of each seed batch, which was used to establish discriminant models through various algorithms. Several preprocessing methods based on hyperspectral data were compared, including multiplicative scatter correction (MSC), standard normal variable (SNV), and first derivative (FD). Then selection methods for identifying informative features in the above data were compared, including successive projections algorithm (SPA), uninformative variable elimination (UVE), and competitive adaptive reweighted sampling (CARS). The results showed that support vector machine (SVM) modeling of machine vision data could distinguish Astragalus seeds from SM with >99% accuracy, but could not satisfactorily distinguish AMM seeds from AM. The FD-UVE-SVM model based on hyperspectral data reached 100.0% accuracy in the validation set. Another 90 seeds were tested, and the recognition accuracy was 100.0%, supporting the stability of the model. In summary, HSI data can be applied to

discriminate among the seeds of AMM, AM, and SM non-destructively and with high accuracy, which can drive standardization in the Astragalus production industry.

KEYWORDS

Astragalus seeds, similar seeds, classification, machine vision, hyperspectral imaging

1 Introduction

Astragalus commonly refers to *Astragalus membranaceus* var. *mongholicus* (AMM) or *A. membranaceus* (AM), a legume which is harvested for the medicinal properties of its roots, which are purported to confer anti-inflammatory, anti-oxidative, and anti-cancer effects (Bai et al., 2018; Chen et al., 2020; Zhang et al., 2021). Notably, Astragalus is among the 40 most commonly used traditional Chinese medicines, and its demand is continuously increasing.

However, the seed quality of traditional Chinese medicines is often markedly lower than that of other crops, with problems including seeds of mixed and unknown origins, contamination with debris and foreign matter, inconsistent maturity, low germination rates, and slow or irregular emergence. The adulteration of traditional Chinese medicines is common due to the high value and high demand for these products, which has aroused widespread concerns for public safety (Reid et al., 2006; Yang et al., 2021b). Because it is primarily propagated by seed, Astragalus production faces particular challenges, such as mixed germplasm resources, uneven seed quality, and substantial adulteration of commercial seeds. These cumulative factors adversely affect Astragalus cultivation, and present the challenge of distinguishing between Astragalus and similar (SM) seeds.

AMM and AM are both included in the 2020 edition of the Chinese Pharmacopoeia, and although *A. complanatus*, *Melilotus officinalis*, *A. sinicus*, and *Hedysarum polybotrys* are also considered authentic medicines, they are common SM seeds found as contaminants in AM or AMM seed lots. In addition to contamination with SM seeds, AM and AMM seeds are often mixed, which is problematic due to differences in growth habit, planting adaptability, cultivation and management techniques, chemical composition, and commercial value. AMM plants are morphologically shorter than AM, but the roots are typical taproot, with characteristically few root branches and high yield (since the root is the harvested portion). By contrast, AM plants are taller, with shorter roots that exhibit higher branch number, and lower yield (Wang and Liu, 1996; Zhang et al., 2009; Yang et al., 2020). Therefore, AMM is the most widely cultivated variety, and mixing AMM and AM seeds will lead to different heights during planting, which limits the effectiveness of AMM management practices and decreases yield. This ongoing problem in AMM cultivations points to a need for a system that can accurately sort AMM and AM seeds.

Conventional methods of distinguishing between AMM, AM, and SM seeds include observations with an electron microscope, physical and chemical methods, ultraviolet spectroscopy, and molecular labeling (Yan et al., 2001; Yan et al., 2005; Wang et al., 2005; Duan et al., 2012; Zheng et al., 2019). However, these methods generally produce qualitative results that depend heavily on experience, especially visual evaluation, and a non-destructive, accurate, and simple method for discriminating among AMM, AM and SM seeds is urgently needed by Astragalus producers and market regulators alike.

Machine vision technology combines computational analysis with image recognition and processing technologies (Savakar and Anami, 2009; Patel et al., 2012). Machine vision with image processing (typically RGB images) is currently widely used in agriculture, while research on seed classification is also developing (Cheng et al., 2010; Huang and Cheng, 2017; Tu et al., 2021). Although machine vision can identify morphological and textural variation well, it does not provide any non-visual trait data, such as internal composition, which has limited the application of this approach for distinguishing seeds from genetically and phenotypically similar varieties. This problem has been overcome through hyperspectral imaging (HSI), which simultaneously integrates spatial data with spectral information to highlight differences in chemical composition that affect light reflection or transmittance through the sample, providing additional analytical layers for each sample. In addition, HSI provides high spatial resolution, generating continuous and narrow-band spectral information

Abbreviations: AMM, *Astragalus membranaceus* var. *mongholicus*; AM, *A. membranaceus*; SM, Similar; SVM, Support vector machine; HSI, Hyperspectral imaging; MSC, Multiplicative scatter correction; SNV, Standard normal variable; FD, First derivative; SPA, Successive projections algorithm; UVE, Uninformative variable elimination; CARS, Competitive adaptive reweighted sampling; PLS-DA, Partial least squares discriminant analysis; MLP, Multilayer perceptron; VIS/NIR, Visible/near-infrared; ROI, region of interest; EWs, Effective Wavelengths; RMSE, The root means square error; LVs, latent variables.

for a given object (Guo et al., 2017; Paoletti et al., 2019). HSI techniques are increasingly tested for application in the identification of crops such as staple grains, fruits, and vegetables (Choudhary et al., 2009; Nansen et al., 2015; Sun et al., 2016; Wang et al., 2016; Wang et al., 2018; Xiao et al., 2020; Liu et al., 2022), supporting its feasibility for distinguishing different seed types.

It should be noted that analyzing the large datasets obtained by machine vision or HSI have presented a non-trivial challenge for data scientists and researchers. Machine learning methods have been developed that are currently the most efficient approaches for image processing and analysis. Common machine learning algorithms, including support vector machine (SVM), partial least squares discriminant analysis (PLS-DA), and multilayer perceptron (MLP), have been successfully applied to a range of classification tasks (Yang et al., 2015a; Sun et al., 2016; Sun et al., 2021; Nazari et al., 2021).

Machine vision and HSI approaches have been combined with machine learning algorithms to classify different crop seeds (Hong et al., 2016; Sun et al., 2017a; Zhao et al., 2018; Nie et al., 2019; Nazari et al., 2021; Ruslan et al., 2022; Tu et al., 2022). Table 1 shows details of these classification tasks. For Astragalus, Xiao et al. (2020) used visible/short-wave near-infrared and near infrared hyperspectral imaging with a convolutional neural network to identify Radix Astragali from five geographical origins, showing an accuracy of >98%. Despite the success of machine vision and HSI technologies in different crop seeds, to our knowledge, no studies have yet reported the application of machine vision and HSI technologies in AMM, AM and SM seed identification.

In this study, machine vision and HSI techniques were applied to the identification of AMM, AM and SM seeds of different origins. The specific objectives of this work include: (1) to establish detection models for AMM, AM and SM seeds using machine vision or HSI image data combined with a machine learning algorithm; (2) to determine the optimal classification model based on the predictive accuracy of different models with the validation dataset; (3) to compare the detection accuracy of the strongest machine vision- or HSI-based models for sample seeds not included in the training data; (4) to compare machine vision and HSI methods to determine which imaging method is better suited for AMM, AM and SM seed identification.

2 Materials and methods

2.1 Materials

The seed samples used in this study were divided into "Astragalus seeds" and "SM seeds". Astragalus seeds include AMM and AM seeds, while SM includes *Astragalus complanatus*, *Melilotus officinalis*, *A. sinicus*, and *Hedysarum polybotrys* seeds. After the seeds were collected, they were sealed and fumigated with aluminum phosphide for 4-5 days, placed in a ventilated place for 7-10 days, bagged, and stored in a ventilated place at room temperature. All samples were identified by field planting and their known origin. The source and quantity of seed samples collected in this study are shown in Table 2.

2.2 RGB image acquisition and feature extraction

Several AMM, AM, and SM seeds were randomly selected and scanned with a ScanMaker i360/i460 scanner (Shanghai, China) at a resolution of 600dpi. The images were saved in TIFF lossless format. Hundreds of seeds were scanned each time, though there was no contact between seeds.

The Phenoseed automatic extraction system, which was jointly developed by the Seed Science and Technology Research Center of China Agricultural University (Beijing, China) and Nanjing Zhinong Yunxin Big Data Technology Co., Ltd. (Nanjing, China), was used to extract phenotypic features of the seeds. Shape features included length (mm), width (mm), L/W ratio, area (mm2), perimeter (mm), and roundness (mm). Color features included R (red in the primary color light spectrum), G (green in the primary color light spectrum), B (blue in the primary color light spectrum), L (luminosity), a (range from red to green), b (range from blue to yellow), hue, saturation, value, gray, and standard deviation. Texture features included the average value and standard deviation of contrast, dissimilarity, homogeneity, energy, correlation, ASM, and entropy under gray, R, G, and B. There were a total of 54 phenotypic features.

2.3 Hyperspectral reflectance data extraction

2.3.1 HSI system and analysis software

Hyperspectral reflectance images of seeds were collected using a prototype visible/near-infrared (VIS/NIR) HSI system installed at the Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University. The spectral range of the system was 311- 1090 nm, the bandwidth was 0.78 nm, and the image resolution was 1004 × 1002 pixels. The software for collecting spectral information was Spectral Image-VNIR (Isuzu Optics Corp., Taiwan, China). Other detailed descriptions of the VIS/NIR HSI system are listed in a paper published by Zhang et al. (2020).

2.3.2 Hyperspectral image acquisition and data extraction

Before collecting hyperspectral images, we calibrated the black-and-white board to produce the corrected image and then

TABLE 1 Application of machine vision and HSI for different crop seed classification tasks.

Method	Seed	Varieties	Classifier(s)	Result	Reference
Machine vision	Sorghum	10 sorghum cultivars	MLP, MLT	99%	Nazari et al., 2021
	Weedy rice	5 cultivated rice varieties and a weedy rice seed	LR	92.40%	Ruslan et al., 2022
	Rice	6 common cultivated rice seed varieties	KNN, SVM, RF	90.54%	Hong et al., 2016
HSI	Grape	3 grape seed varieties	SVM	94.30%	Zhao et al., 2018
	Maize	Jingke 968 and Non-Jingke 968	RF, SVM, MLP	~99%	Tu et al., 2022
	Hybrid okra seeds and hybrid loofah seeds	6 hybrid okra seed varieties and 6 hybrid loofah seed varieties	PLS-DA, SVM, DCNN	~95%	Nie et al., 2019
	Rice	Rice seeds from 4 different regions	SVM	91.67%	Sun et al., 2017a

set the moving speed of the electric control platform to 1.7 mm/s and the exposure time of the camera to 6 ms. When collecting spectral information from the sample, we placed each seed on the electric control displacement table. The camera scanned the whole platform as it moved. We then collected 50 seeds from each batch, for a total of 1150 seeds. The hyperspectral image acquisition process was completed in the dark box.

Before extracting spectral information, an HSI analyzer (Isuzu Optics Corp., Taiwan, China) was used to correct the spectral image, as shown in equation (1), where I is the corrected hyperspectral image. I_0 is the original hyperspectral image; B is the blackboard file image (reflectivity close to 0%), and W is the whiteboard file image (reflectivity close to 100%).

$$I = \frac{I_0 - B}{W - B} \quad (1)$$

After the black-and-white plate was corrected, the single seed was separated from the background of the hyperspectral image. The background and seeds were separated by setting a threshold and obtaining the region of interest (ROI) by

TABLE 2 Table of materials.

Category	Species	Identifier	Source	Genotypes	Number of seeds	
					Machine vision	Hyperspectral
Astragalus seeds	*A. membranaceus* var. *mongholicus* (AMM)	AMM-1	Tangshan, Hebei	Cultivated	234	50
		AMM-2	Datong, Shanxi	Cultivated	249	50
		AMM-3	Chifeng, Inner Mongolia	Cultivated	199	50
		AMM-4	Yulin, Shaanxi	Cultivated	215	50
		AMM-5	Guyuan, Ningxia	Cultivated	184	50
		AMM-6	Shangzhuang, Beijing	Cultivated	361	50
		AMM-7	Anguo, Hebei	Cultivated	85	50
		AMM-8	Market purchase	Cultivated	186	50
		AMM-9	Longnan, Gansu	Wild	234	50
	A. membranaceus (AM)	AM-1	Tangshan, Hebei	Cultivated	290	50
		AM-2	Zhangjiakou, Hebei	Cultivated	256	50
		AM-3	Hezheng, Gansu	Cultivated	237	50
		AM-4	Guyuan, Ningxia	Cultivated	197	50
		AM-5	Shangzhuang, Beijing	Cultivated	239	50
		AM-6	Longnan, Gansu	Wild	202	50
		AM-7	Zhangjiakou, Hebei	Wild	240	50
Similar (SM) seeds	*A. complanatus*	AC-1	Anguo, Hebei	Cultivated	427	50
		AC-2	Bozhou, Anhui	Cultivated	325	50
		AC-3	Suqian, Jiangsu	Cultivated	407	50
		AC-4	Laboratory samples	/	352	50
	M. officinalis	MO-1	Laboratory samples	/	132	50
	A. sinicus	AS-1	Laboratory samples	/	214	50
	H. polybotrys	HP-1	Laboratory samples	/	145	50

morphological filtering and mask processing. Due to the large noise interference in the head and tail bands, only the reflection spectra of 765 bands within 400-1000 nm of each seed were extracted for subsequent modeling and analysis. Among them, the VIS included 490 reflectance data points in the range of 400-780 nm, and the NIR included 275 reflectance data points in the range of 780-1000 nm.

2.4 Spectra preprocessing

Spectral information is inevitably affected by various factors, and spectral preprocessing is used to improve the usefulness of spectral data (Esquerre et al., 2012). This study compared three spectral preprocessing methods, including multiplicative scatter correction (MSC), standard normal variable (SNV), and first derivative (FD). From the latter, MSC and SNV were used to consider the addition/multiplication effect and scattering effect in spectral data (Silalahi et al., 2018; Wu et al., 2019). FD helps delete baseline offset (Qu and Liu, 2017). These preprocessing methods eliminate the external interference generated during the acquisition of hyperspectral images. In this study, a suitable preprocessing method was selected based on the preprocessing effect of various preprocessing methods on the raw spectral data.

2.5 Selection of effective wavelengths

The presence of high dimensional data and a large amount of redundant information in hyperspectral images can affect the modelling speed. Therefore, it is important to use the variable selection method during the analysis and processing of hyperspectral data. In this study, successive projection algorithms (SPA), uninformative variable elimination (UVE), and competitive adaptive reweighted sampling (CARS) were used to select EWs.

SPA can extract low collinearity and low redundancy variables to avoid the influence of information overlap and collinearity. When the SPA method is used to optimize the wave band, multiple linear regression models can be established one by one for different wave segment subsets, and the root means square error (RMSE) value can be calculated, in which the number of variables corresponding to the lowest RMSE is the optimal EWs (Galvão et al., 2008; Zhang et al., 2018b). UVE can remove the wavelength variables that contribute less to the modeling and select the characteristic wavelength variables. The removed wavelength variables are called non-information variables. UVE and CARS establishment are based on the PLS algorithm. To select non-informative variables, the UVE algorithm adds a group of white noise variables with the same number of original variables to the PLS model and obtains the regression coefficient corresponding to each variable based on the cross-leave method of the PLS model. The stable value of each variable coefficient is divided by the standard deviation, their quotient is compared with the stable value obtained from the random variable matrix, and those wavelength variables that are invalid for modeling are deleted as random variables. (Put et al., 2006; Wang et al., 2020) In CARS algorithm, adaptive weighted sampling (ARS) is used to retain the points with large absolute values of regression coefficients in the PLS model as a new subset each time, and remove the points with small weights. Then, the PLS model is established based on the new subset. After multiple calculations, the subset with the smallest cross-validation root mean square error (RMSECV) of the PLS model is selected as the characteristic wavelength. (Zhang et al., 2018a; Zhou et al., 2020a)

2.6 Data–driven modeling

Support vector machine (SVM), partial least squares discriminant analysis (PLS-DA), and multilayer perceptron (MLP) are widely used classification methods that have been well validated for seed detection (Kujawa et al., 2014; Nazari et al., 2021; Yang et al., 2021a). In this study, three classification models of AMM, AM, and SM seeds were established using SVM, PLS-DA, and MLP, respectively. SVM is widely used to solve linearly differentiable and linearly indistinguishable classification problems, and the radial basis kernel function (RBF) kernel is the most common and effective method for classification problems. Optimization of the hyperparameters is necessary before the real model can be trained. In current practical applications, hyperparameters are usually determined empirically or by grid search. (Liu et al., 2022) PLS-DA is a typical classification method, which is considered as a supervision method to distinguish samples to the maximum extent. (Nie et al., 2019; Zhang et al., 2022b) MLP is a feedforward neural network. It maps a set of input vectors to a set of output vectors. The inputs and outputs can be connected by multilayer weighting, with strong self-learning, adaptive, associative memory and parallel processing of things and environments. (Xu et al., 2021)

To avoid the effect of default parameters on the prediction accuracy of the classification model, the internal parameters of the classification model must be separately adjusted. In the SVM algorithm, the RBF kernel was selected, and it carried out the 5-fold internal cross-validation and grid search method to calculate optimal penalty coefficient c and the kernel parameter g. The searching range was both set to -10 to 10 with the step of 0.2 (a total of 101*101 combinations were used to search the best parameters). In the PLS-DA model, the number of latent variables (LVs) changes, and the model correctly identifies the highest percentage of seeds. The MLP network with two hidden layers was selected, and the hidden layer adopted the hyperbolic tangent activation function of SPSS. The output layer adopted the Softmax activation function.

2.7 Analyzing

MSC, SNV, FD spectral preprocessing, SPA, UVE, CARS feature band extraction, and SVM and PLS-DA model driving processes were implemented in Matlab R2020b. The MLP modeling process was using IBM SPSS statistics 26. For each model, the ratio of three model training sets and validation sets was 7:3. The specific quantity of seeds in each batch is shown in Table 2. Origin 2022 and R 4.1.2 were used to visualize data. The experimental flow of this study is shown in Figure 1.

3 Results

3.1 Identification of AMM, AM, and SM seeds based on machine vision

In order to develop a machine learning algorithm capable of distinguishing AMM, AM, and other SM seeds, we first tested untrained machine vision using a mixed set of 5610 seeds. The machine vision technology based on Phenoseed automatic extraction system extracted a set of 54 potentially informative morphological features for discriminating among seed types, including shape, color, and texture. A probability density map was then generated to examine the distributions of these phenotypic features across the 5610 combined AMM, AM and SM seeds (Figure 2), which showed high overlap in their features related to color and texture, especially between AMM and AM. These results indicated that these features might not be sufficiently different between seed types to form a basis for distinguishing between them. It is worth noting that the size of SM was generally smaller than that of AMM or AM seeds, although the size distributions showed substantial overlap (Figure 2). Thus, SM resembled small AMM and AM seeds, which is germane to distinguishing SM seeds mixed into Astragalus seed lots in the market.

In order to establish a model for classifying AMM, AM, and SM seeds, the 54 morphological features detected by machine vision were used as inputs for SVM, PLS-DA, and MLP networks, with 3927 seeds in the training set and 1683 seeds in the validation set. The accuracy in distinguishing between seeds was then calculated for each model (Figure 3), with 48 latent variables (LVs) selected for the PLS-DA model. Classification accuracy of SM seeds using these features reached >98.2%, but the classification effect of AMM and AM seeds were not ideal. In particular, SVM showed 83.3% accuracy for identifying AMM seeds, 89.2% accuracy for detecting AM seeds, and 99.5% accuracy for SM seeds, resulting in 91.1% average accuracy for this model.

3.2 Identification of AMM, AM, and SM seeds using HSI

Since machine vision with different machine learning algorithms could not effectively distinguish AMM seeds from AM seeds, we next explored means of improving accuracy by VIS/NIR hyperspectral imaging of reflectance spectra between 400 nm and 1000 nm bands for each seed type.

3.2.1 Spectral characteristics

To identify wavelengths that were distinct among seed types, the reflectance spectra were obtained for 1150 seeds (Figure 4A) and the average spectrum of AMM, AM, and SM seeds were calculated (Figure 4B), which revealed that the average spectra of AMM and AM seeds was significantly lower than that of SM seeds. By contrast, AMM and AM seeds displayed similar spectra, which agreed well with previous studies examining the variety and viability of seeds from other crops (Yang et al., 2017; Wakholi et al., 2018; Yang et al., 2021a). However, the reflectance of AMM was generally higher than that of AM seeds, with greater differences in the NIR region (780-1000

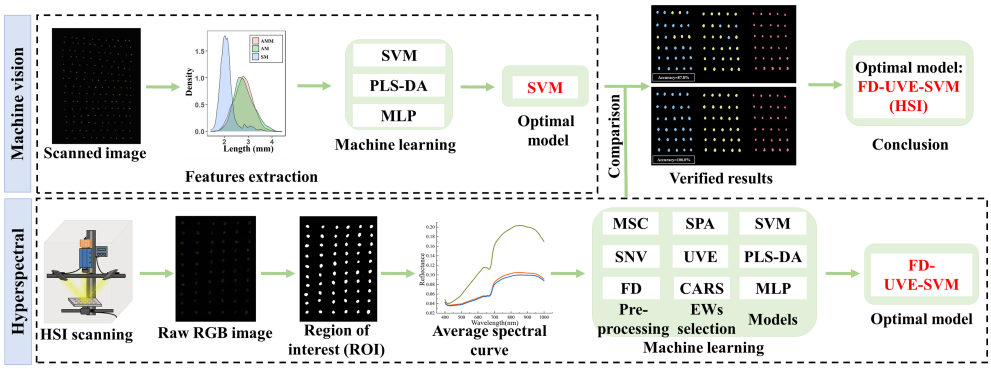


FIGURE 1
Technical routes.

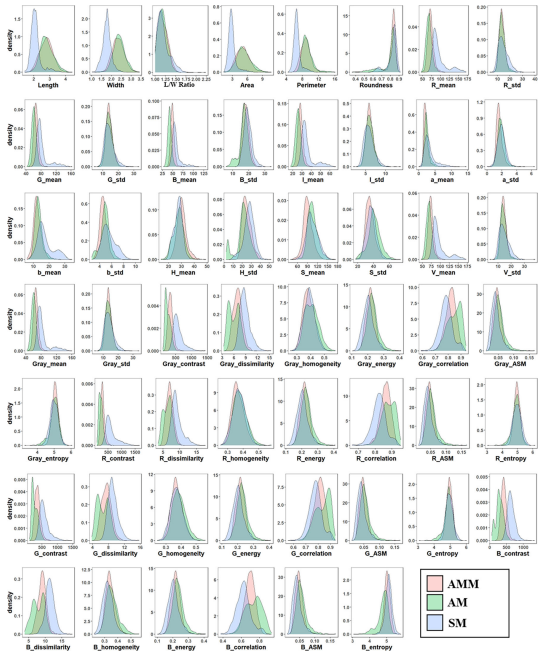


FIGURE 2
Probability density distribution of phenotypic features of AMM, AM, and SM seeds.

nm) than in the VIS region (400-780 nm). The VIS region may be related to β-carotene and anthocyanin in the seeds (Sun et al., 2021), and the difference in the NIR region may be related to protein, starch and other organic matter in the seeds (Cen and He, 2007; Awanthi et al., 2019). These results indicated that HSI could capture differences in texture, pigment and other physical and chemical properties between AMM and AM seeds. And the differences between AMM and AM seeds in terms of organic matter such as protein and starch are greater than the differences in pigment content, which explains the inability of models based on VIS data (machine vision) to correctly distinguish between AMM and AM. Further analysis showed that in the range of 660-750 nm, the spectral curve showed an obvious upward trend, and the average spectral curve of the two seeds gradually separated. The reason is that this wavelength corresponds to the vibration of N-H chemical bond of amino acid in seeds, which can be used to verify the difference of amino acid content in AMM and AM seeds. (Yang et al., 2021a; Wang et al., 2022) In addition, there are four absorption peaks in the average reflectance spectra of AMM and AM seeds (valleys at 415 nm, 640, 680 and 885 nm). The carotenoid (Yang et al., 2021a) and proanthocyanidin content of the seed coat (Wang et al., 2022) can be determined at about 415 nm; The bands at about 640 nm and 680 nm may be associated with the absorption of chlorophyll b and chlorophyll a (Zhang et al., 2016). The band at about 885 nm may be associated with C-O, N-H, C-H and O-H bonding vibrations in proteins, carbohydrates and fats (Caporaso et al., 2018). However, the spectral curves were not unique enough to distinguish AMM, AM, and SM seeds. More specifically, while a large proportion of SM overlapped with Astragalus seeds in the original spectral curve, the reflectance spectra of AMM and AM showed extremely high overlap, making them indistinguishable by curve shape. To identify specific spectral features or effective wavelengths that could be used for classification of AMM, AM, or SM seeds, it was first necessary to generate discriminant analytical models to test different methods of preprocessing to maximize the accuracy of discriminating among AMM, AM and SM seeds.

3.2.2 Spectral preprocessing and effect analysis

In order to establish the most effective discriminant model, we first tested three methods for preprocessing the seed reflectance spectra, including MSC, SNV and FD, as well as raw data. The SVM, PLS-DA, and MLP methods were each used to generate models with the processed and raw data, using a training set of 805 seeds and a validation set of 345 seeds (i.e., a 7:3 ratio) (see

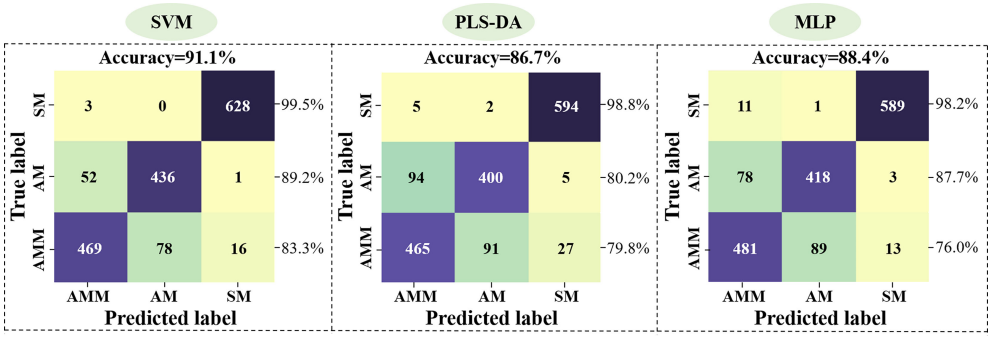


FIGURE 3
Performance of the SVM, PLS-DA, and MLP model validation sets based on machine vision data.

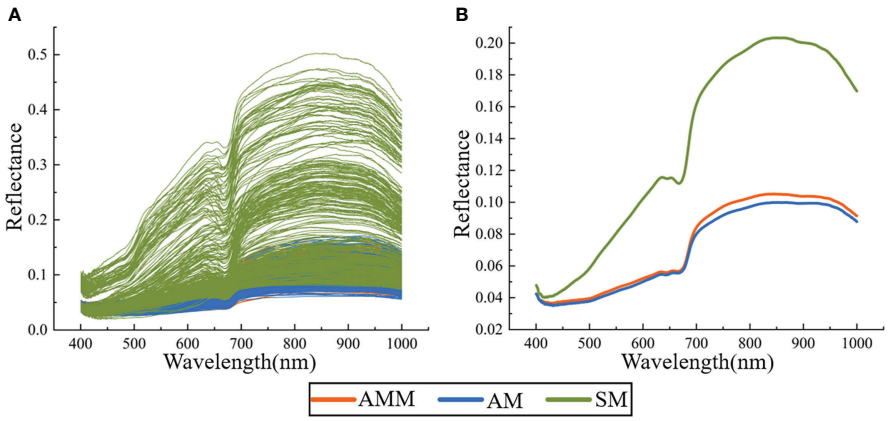


FIGURE 4
Reflection spectrum **(A)** and average spectrum **(B)** of AMM, AM, and SM seeds.

section 2.6 for parameter details). The PLS-DA model selected 10, 10, 10, and 9 LVs from raw spectra or spectra preprocessed by MSC, SNV, FD, respectively. Validation of each classifier model with each respective preprocessing method showed that accuracy ranged from 81.7%, in the PLS-DA model with raw data, to 100.0% in the SVM model with the FD-processed spectra (Figure 5). Among the three pre-processing methods, the model built after FD pre-processing was the optimal, while MSC and SNV showed suboptimal accuracy compared to FD, but still better than RAW. Among the three classifiers, the performance of the model showed SVM > MLP > PLS-DA, regardless of the pre-processing based. While accuracy differed to a limited extent between each combination of classifier and data processing method, it warrants mention that all preprocessing methods increased the accuracy over that of raw data input for all classifiers by denoising the original reflectance spectra. Since the FD-SVM denoising/classifier combination provided the highest accuracy in discriminating among AMM, AM, and SM seeds, this method was used in subsequent analyses of effective wavelengths HSI reflectance data.

3.2.3 Selection of effective wavelengths

Based on the above results showing highly accurate classification of AMM, AM, and SM seeds obtained by full

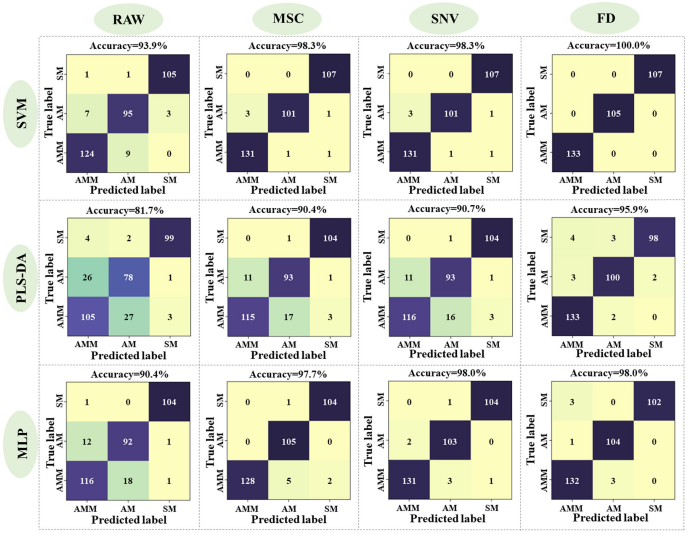


FIGURE 5
Accuracy of SVM, PLS-DA, and MLP models with spectral data that was preprocessed or not.

spectra models, we next sought to increase computational efficiency and reduce processing time by identifying effective wavelengths in FD-processed data that were informative for classifying seed types.

Since the algorithmic principles underlying wavelength selection can impact modeling results (Zhang et al., 2020), we tested three approaches to EW extraction, SPA, UVE, and CARS. In SPA extraction, the root means square error (RMSE) values decreased with increasing number of variables (i.e., EWs). The lowest RMSE value coincided with 22 variables, above which RMSE remained stable (Figure 6A). Therefore, an RMSE threshold of 0.5164 with 22 effective wavelengths in the denoised reflectance spectrum was selected for subsequent tests (Figure 6B). By contrast, UVE analysis identified 391 potentially informative EW features (Figure 6C) for classifier analysis. In CARS extraction, the number of EWs decreased rapidly with the exponential decay function, but decreased at a slower rate as the number of samples increases (Figure 6D). The cross-validation root mean square error (RMSECV) tended to decrease and then increase as the number of samples increased, with the smallest RMSECV value when the number of samples reached 14 (Figure 6E) (i.e., the subset of EW selected for this sampling was the key variable for predicting AMM, AM and SM seeds). Therefore, 140 potentially informative EW features (Figure 6D) were identified for classifier analysis after CARS.

The EWs selected by SPA were further examined to better understand their relevance as a theoretical basis for distinguishing AMM, AM and SM seeds. Several EWs were located around specific regions, which suggested that seed reflectance was related to the presence of specific chemicals. Some specific wavelengths in the visible light spectrum are reportedly related to plant pigments, such as absorption peaks for carotenoids at 412.9 and 413.7 nm, chlorophyll a at 427.8, 430.7, and 666.4nm, and anthocyanin at around 435.2 and 437.5 nm (Li et al., 2013; Nansen et al., 2015; Yang et al., 2015b; Zhang et al., 2016). Among the EWs screened by SPA, the selection of bands near pigment absorption peaks indicated that accurate classification of AMM, AM, and SM seeds may rely on seed color. By contrast, several EWs in the NIR were attributable to various chemical bonds. For example, EWs located near 833.1 nm were related to the vibrations of C-O, N-H, C-H, and O-H bonds in proteins, carbohydrates, and fats, respectively (Cen and He, 2007; Sun et al., 2017b; Caporaso et al., 2018). Tannin absorption peaks were also detected near 887.7 and 898.1 nm, and cellulose absorption peaks were present near 935.0 nm (Wang et al., 2022). In addition, peaks between 860-970 nm reflected the vibration of N-H bonds in proteins and amino acids, and suggested that these seeds differed in protein or amino acid contents (Zhou et al., 2020b). Since tannin, cellulose, and protein contents in the seed coat affect the thickness and hardness of the seed coat, these results indicated that AM, AMM, and SM seeds differed in their seedcoat hardness.

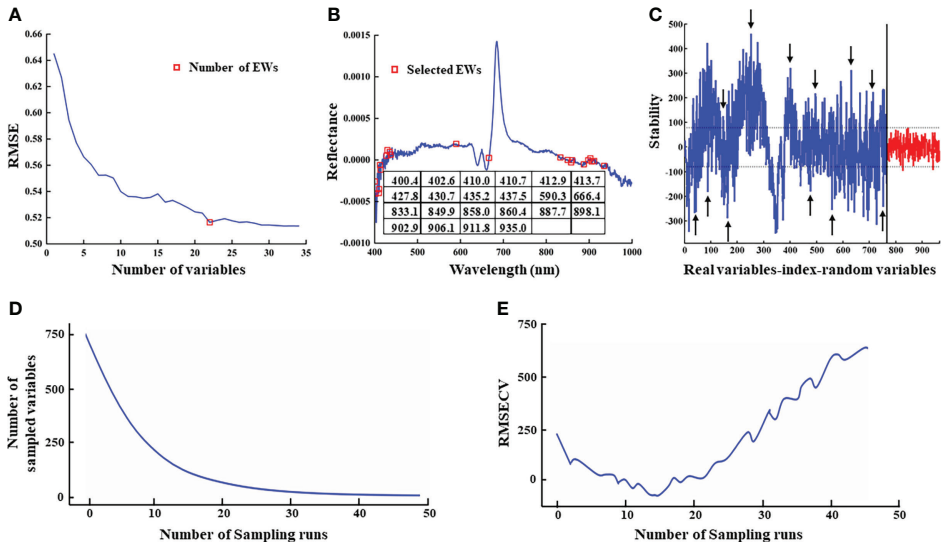


FIGURE 6
Results of EWs selection. **(A)** RMSE for the number of EWs by SPA; **(B)** The position of EWs identified by SPA in FD-processed spectrum; **(C)** EWs selected by UVE. Peaks in the blue curve represent stability values for the 765 wavelengths; the red curve shows the distribution of stability values of random noise variables generated by UVE which were used to determine the thresholds. The two horizontal dashed lines indicate the upper and lower thresholds of band selection. Bands between the thresholds were eliminated as invalid information, and the remaining variables were selected. The arrowheads show a few representative wavelengths; **(D)**Variation of wavelength variables number by CARS; **(E)** Variation of RMSECV by CARS.

3.2.4 SVM modeling based on different EW selection methods

In light of our above findings, we next tested the accuracy of the SVM model in discriminating between AM, AMM, and SM seeds using FD-denoised EWs selected by SPA or UVE or CARS as inputs. Validation of model accuracy is shown in Figure 7. The changes in model performance indicate that UVE screening of EWs results in higher accuracy, with CARS showing the sub-optimal accuracy and SPA the worst performance. It potentially due to an insufficiently large sample set of EWs selected by SPA, leading to the exclusion of informative wavelengths related to seed classification. Our results thus showed that UVE method is more suitable than the SPA and CARS algorithm for building SVM models to classify AMM, AM, and SM seeds and that the FD-UVE-SVM combination provides an optimal model for this task, providing 100.0% accuracy in the validation set.

3.3 Visual comparison between machine vision and HSI prediction results

Based on the above results showing that SVM classification provided the highest accuracy with machine vision imaging data, while the FD-UVE-SVM model showed the highest accuracy with HSI-based data, we next visually examined a set of 90 seeds, including 30 AMM, 30 AM, and 30 SM, to manually verify the classifications made by these two models (Figure 8). The SVM model based on machine vision image data (Figure 8A) could correctly predict the 30 SM seeds, while 7 of the 30 AMM seeds were incorrectly predicted as AM seeds, and 4 of the 30 AM seeds were incorrectly predicted as AMM seeds (i.e., an overall accuracy of 87.8%), indicating sub-optimal discriminatory power between AM and AMM seeds. By contrast, FD-UVE-SVM modelling of HSI data correctly predicted the types of all 90 seeds (i.e., 100.0% accuracy), indicating that hyperspectral data contained sufficient information to accurately discriminate between AMM, AM, and SM seeds, supporting its application in routine analysis required for seed sorting and quality control in production settings.

4 Discussion

In this study, HSI data outperformed machine vision image data in SVM-based models for distinguishing AMM, AM and SM seeds, which aligns well with studies comparing these approaches for classifying kernels of rice (Fabiyi et al., 2020) and maize (Tu et al., 2022), as well as comparing chlorophyll content in sorghum leaves (Zhang et al., 2022a). It is likely that machine vision data resulted in lower accuracy in classifying AMM and AM seeds because the features extracted by machine vision were purely morphological phenotypes. Since AMM and AM belong to the same species, they share highly similar seed morphology. Thus, other phenotypic data, such seed traits related to the internal accumulation or deposition of specific metabolites affecting light reflection or diffraction captured by HSI data can provide more informative features for discriminating among subspecies. Although both methods are non-destructive and high-throughput tests that enable screening of intact seeds, each technology is accompanied by advantages and disadvantages. The main advantages of machine vision over HSI are the relatively low instrument cost and faster image acquisition, which are linked to its main disadvantage of capturing strictly morphological information that cannot account for many internal, physiological seed traits. By contrast, HSI data includes hundreds or thousands of spectral bands, and therefore contains more information allowing more robust discrimination among samples. However, it should be pointed out that the high dimension of spectral data limits the calculation speed and processing time to some extent. Moreover, HSI equipment is relatively expensive and the operating cost is high.

In HSI data analysis, FD pre-processing achieved superior results to MSC and SNV, which aligns well with studies for classifying kernels of sugar beet (Yang et al., 2018) and the

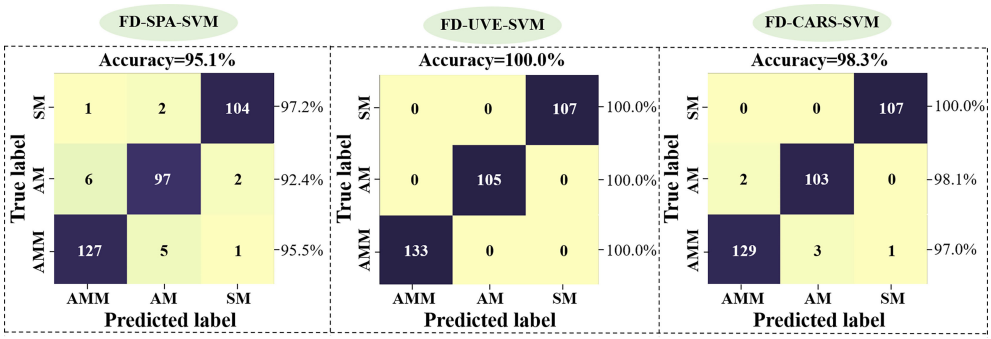


FIGURE 7
Validation of SVM model accuracy based on EW data selected by SPA (left), UVE (middle) or CARS (right) methods.

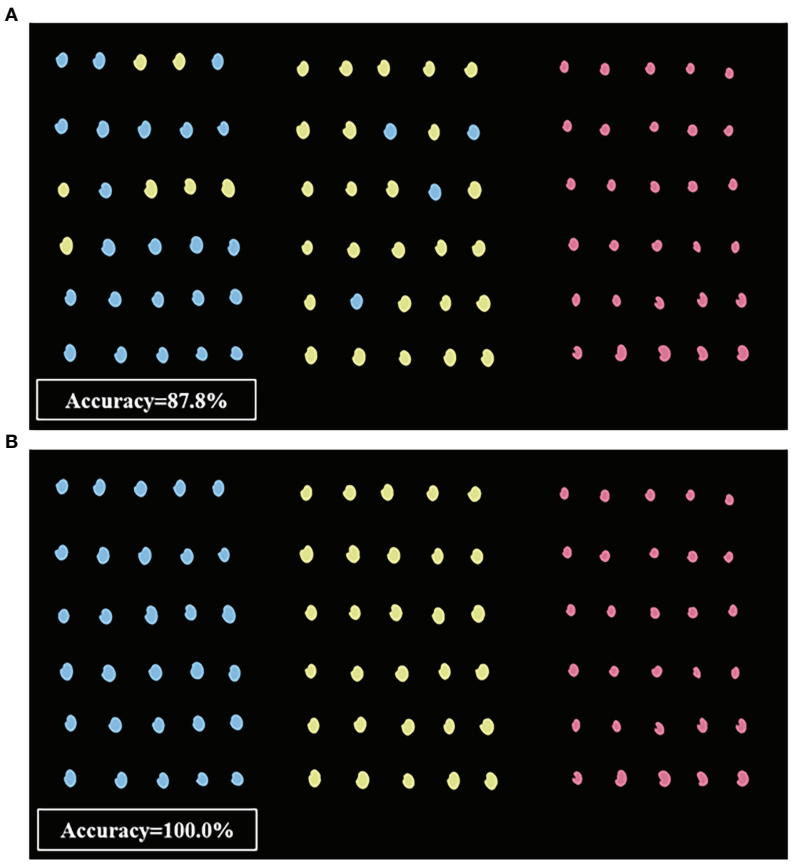


FIGURE 8
Visualization results based on **(A)** machine vision optimal model and **(B)** HSI optimal model (from left to right: AMM, AM, and SM seeds).

detection of germination rates in sorghum-sudan grass seed (Hui et al., 2022). It is likely that due to FD eliminates baseline drift in AMM, AM and SM seed spectral data and improves the spectral band characteristics and spectral resolution, thus FD pre-processing achieves the highest accuracy (Qu and Liu, 2017; Li et al., 2021). However, it is worth noting that spectral derivatives can also increase the noise level and reduce the spectral signal-to-noise ratio, which is detrimental to modeling. The higher the derivative order, the more serious the degradation of the signal-to-noise ratio. Therefore, in spectral analysis, only FD or second derivative (SD) is generally used for spectra.

Selecting a subset of EW features by SPA or UVE or CARS to construct the model can dramatically reduce processing time (Gao et al., 2013). SPA-based screening of denoised EWs captures a portion of bands in the NIR wavelength range (>780nm), whereas machine vision technology collects wavelengths largely in the VIS range. These additional NIR EWs may explain why HSI technology can provide higher accuracy than machine vision in distinguishing AMM, AM, and SM seeds.

Although HSI technology has enormous potential for discrimination among highly (visually) similar seeds, sample sizes in this study, including nine batches of AMM, seven batches of AM, and seven batches of SM, were not adequately large for robust statistical verification. Follow-up research will necessarily include larger sample sets improve the generalizability and accuracy of the classification model. In addition, in order to maintain model validity when testing seeds from different origins and growing seasons, the model can be updated using a method established for updating a maize seed detection model (Guo et al., 2017; Tu et al., 2022).

5 Conclusion

We compared machine vision and HSI image data to classify AMM, AM, and SM seeds, which led to the following specific conclusions:

1) SVM-based models of machine vision image data to distinguish AMM, AM, and SM seeds indicated that

AM/AMM seeds could be distinguished from SM seeds at >99.0%, but could not well-distinguish between AMM and AM seeds.

2) The application of a FD-UVE-SVM model to HSI data resulted in 100.0% accuracy, thus validating the SVM classification model as the best suited for distinguishing SM, AM, and AMM seeds.

3) Verification of model accuracy based on machine vision and HSI data from a 90-seed verification set indicated that predictive accuracy was 100.0% with HSI data, demonstrating the efficiency, reliability, and simplicity of this model, and importantly, revealing that HSI is more suitable for discriminating among AMM, AM, and SM seeds.

In general, this study used machine vision and HSI technology to classify AMM, AM, and SM seeds. FD-UVE-SVM modeling of HSI data can be used to accurately distinguish AMM, AM, and SM. This strategy can be adapted for routine analyses in production facilities. These advances can in turn increase the economic benefits of Astragalus seeds.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

QS, XD, and YX designed the experiments. YX collected the RGB, HSI images, analyzed the data, and wrote the manuscript. WW, YC, TZ, KT, YH, and HC provided comments and suggestions for the manuscript and checked the writing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Datong Municipal Government [grant number 201904710111639]; and the National Administration of Traditional Chinese Medicine State Administration of Traditional Chinese Medicine [grant number 202004610111024].

Conflict of interest

Author HC was employed by Hengde Materia Medica Beijing Agricultural Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1031849/full#supplementary-material

References

Awanthi, M. G. G., Jinendra, B. M. S., Navaratne, S. B., and Navaratne, C. M. (2019). Adaptation of visible and short wave near infrared (VIS-SW-NIR) common PLS model for quantifying paddy hardness. *J. Cereal Sci.* 89, 102795. doi: 10.1016/j.jcs.2019.102795

Bai, C., Hao, J., Hao, X., and Feng, M. (2018). Preparation of astragalus membranaceus lectin and evaluation of its biological function. *BioMed. Rep.* 9, 345–349. doi: 10.3892/br.2018.1132

Caporaso, N., Whitworth, M. B., Grebby, S., and Fisk, I. D. (2018). Non-destructive analysis of sucrose, caffeine and trigonelline on single green coffee beans by hyperspectral imaging. *Food Res. Int.* 106, 193–203. doi: 10.1016/j.foodres.2017.12.031

Cen, H., and He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* 18, 72–83. doi: 10.1016/j.tifs.2006.09.003

Cheng, C., Liu, J., Wang, H., and Xiong, W. (2010). Infrared spectroscopic studies of Chinese medicines. *Appl. Spectrosc Rev.* 45, 165–178. doi: 10.1080/05704920903574256

Chen, Z., Liu, L., Gao, C., Chen, W., Vong, C. T., Yao, P., et al. (2020). Astragali radix (Huangqi): A promising edible immunomodulatory herbal medicine. *J. Ethnopharmacol* 258, 112895. doi: 10.1016/j.jep.2020.112895

Choudhary, R., Mahesh, S., Paliwal, J., and Jayas, D. S. (2009). Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosyst. Eng.* 102, 115–127. doi: 10.1016/j.biosystemseng.2008.09.028

Duan, L., Chen, T., Li, M., Chen, M., Zhou, Y., Cui, G., et al. (2012). Use of the metabolomics approach to characterize Chinese medicinal material huangqi. *Mol. Plant* 5, 376–386. doi: 10.1093/mp/ssr093

Esquerre, C., Gowen, A. A., Burger, J., Downey, G., and O'Donnell, C. P. (2012). Suppressing sample morphology effects in near infrared spectral imaging using

chemometric data pre-treatments. *Chemometr. Intell. Lab. Syst.* 117, 129–137. doi: 10.1016/j.chemolab.2012.02.006

Fabiyi, S. D., Vu, H., Tachtatzis, C., Murray, P., Harle, D., Dao, T. K., et al. (2020). Varietal classification of rice seeds using RGB and hyperspectral images. *IEEE Access* 8, 22493–22505. doi: 10.1109/ACCESS.2020.2969847

Galvão, R. K. H., Araújo, M. C. U., Fragoso, W. D., Silva, E. C., José, G. E., Soares, S. F. C., et al. (2008). A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. *Chemometr. Intell. Lab. Syst.* 92, 83–91. doi: 10.1016/j.chemolab.2007.12.004

Gao, J., Li, X., Zhu, F., and He, Y. (2013). Application of hyperspectral imaging technology to discriminate different geographical origins of jatropha curcas l. *seeds. Comput. Electron Agric.* 99, 186–193. doi: 10.1016/j.compag.2013.09.011

Guo, D., Zhu, Q., Huang, M., Guo, Y., and Qin, J. (2017). Model updating for the classification of different varieties of maize seeds from different years by hyperspectral imaging coupled with a pre-labeling method. *Comput. Electron Agric.* 142, 1–8. doi: 10.1016/j.compag.2017.08.015

Hong, P. T. T., Hai, T. T. T., Lan, L. T., Hoang, V. T., Hai, V., and Nguyen, T. T. (2016). "Comparative study on vision based rice seed varieties identification", in *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*. 377–382. doi: 10.1109/KSE.2015.46

Huang, K., and Cheng, J. (2017). A novel auto-sorting system for Chinese cabbage seeds. *Sensors (Switzerland)* 17, 886. doi: 10.3390/s17040886

Hui, Y., Wang, D., Tang, X., Peng, Y., Wang, H., Zhang, H., et al. (2022). Detection of sorghum-Sudan grass seed germination rate based on near infrared spectroscopy. *Spectrosc. Spectral Anal.* 42, 423–427. doi: 10.3964/j.issn.1000-0593(2022)02-0423-05

Kujawa, S., Nowakowski, K., Tomczak, R. J., Dach, J., Boniecki, P., Weres, J., et al. (2014). Neural image analysis for maturity classification of sewage sludge composted with maize straw. *Comput. Electron Agric.* 109, 302–310. doi: 10.1016/j.compag.2014.08.014

Li, J., Huang, W., Zhao, C., and Zhang, B. (2013). A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. *J. Food Eng.* 116, 324–332. doi: 10.1016/j.jfoodeng.2012.11.007

Li, X., Jiang, H., Jiang, X., Shi, M., and Fielke, J. M. (2021). Identification of geographical origin of Chinese chestnuts using hyperspectral imaging with 1D-CNN algorithm. *Agriculture* 11, 1274. doi: 10.3390/agriculture

Liu, Q., Wang, Z., Long, Y., Zhang, C., Fan, S., and Huang, W. (2022). Variety classification of coated maize seeds based on raman hyperspectral imaging. *Spectrochim Acta A Mol. Biomol Spectrosc* 270, 120772. doi: 10.1016/j.saa.2021.120772

Nansen, C., Zhao, G., Dakin, N., Zhao, C., and Turner, S. R. (2015). Using hyperspectral imaging to determine germination of native Australian plant seeds. *J. Photochem. Photobiol. B* 145, 19–24. doi: 10.1016/j.jphotobiol.2015.02.015

Nazari, L., Shaker, M., Karimi, A., and Ropelewska, E. (2021). Identification of sorghum genotypes using a machine vision system. *J. Food Process Eng.* 44, e13673. doi: 10.1111/jfpe.13673

Nie, P., Zhang, J., Feng, X., Yu, C., and He, Y. (2019). Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. *Sens Actuators B Chem.* 296, 126630. doi: 10.1016/j.snb.2019.126630

Paoletti, M. E., Haut, J. M., Plaza, J., and Plaza, A. (2019). Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.* 158, 279–317. doi: 10.1016/j.isprsjprs.2019.09.006

Patel, K. K., Kar, A., Jha, S. N., and Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. *J. Food Sci. Technol.* 49, 123–141. doi: 10.1007/s13197-011-0321-4

Put, R., Daszykowski, M., Baczek, T., and vander Heyden, Y. (2006). Retention prediction of peptides based on uninformative variable elimination by partial least squares. *J. Proteome Res.* 5, 1618–1625. doi: 10.1021/pr0600430

Qu, Y., and Liu, Z. (2017). Dimensionality reduction and derivative spectral feature optimization for hyperspectral target recognition. *Optik (Stuttg)* 130, 1349–1357. doi: 10.1016/j.ijleo.2016.11.143

Reid, L. M., O'Donnell, C. P., and Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends Food Sci. Technol.* 17, 344–353. doi: 10.1016/j.tifs.2006.01.006

Ruslan, R., Khairunniza-Bejo, S., Jahari, M., and Ibrahim, M. F. (2022). Weedy rice classification using image processing and a machine learning approach. *Agric. (Switzerland)* 12, 645. doi: 10.3390/agriculture12050645

Savakar, D. G., and Anami, B. S. (2009). Recognition and classification of food grains, fruits and flowers using machine vision. *Int. J. Food Eng.* 5, 14. doi: 10.2202/1556-3758.1673

Silalahi, D. D., Midi, H., Arasan, J., Mustafa, M. S., and Caliman, J. P. (2018). Robust generalized multiplicative scatter correction algorithm on pretreatment of near infrared spectral data. *Vib Spectrosc* 97, 55–65. doi: 10.1016/j.vibspec.2018.05.002

Sun, J., Jiang, S., Mao, H., Wu, X., and Li, Q. (2016). Classification of black beans using visible and near infrared hyperspectral imaging. *Int. J. Food Prop* 19, 1687–1695. doi: 10.1080/10942912.2015.1055760

Sun, J., Lu, X., Mao, H., Jin, X., and Wu, X. (2017a). A method for rapid identification of rice origin by hyperspectral imaging technology. *J. Food Process Eng.* 40, e12297. doi: 10.1111/jfpe.12297

Sun, J., Zhang, L., Zhou, X., Yao, K., Tian, Y., and Nirere, A. (2021). A method of information fusion for identification of rice seed varieties based on hyperspectral imaging technology. *J. Food Process Eng.* 44, e13797. doi: 10.1111/jfpe.13797

Sun, Y., Liu, Y., Yu, H., Xie, A., Li, X., Yin, Y., et al. (2017b). Non-destructive prediction of moisture content and freezable water content of purple-fleshed sweet potato slices during drying process using hyperspectral imaging technique. *Food Anal. Methods* 10, 1535–1546. doi: 10.1007/s12161-016-0722-0

Tu, K., Wen, S., Cheng, Y., Xu, Y., Pan, T., Hou, H., et al. (2022). A model for genuineness detection in genetically and phenotypically similar maize variety seeds based on hyperspectral imaging and machine learning. *Plant Methods* 18, 81. doi: 10.1186/s13007-022-00918-7

Tu, K., Wen, S., Cheng, Y., Zhang, T., Pan, T., Wang, J., et al. (2021). A non-destructive and highly efficient model for detecting the genuineness of maize variety 'JINGKE 968' using machine vision combined with deep learning. *Comput. Electron Agric.* 182, 106002. doi: 10.1016/j.compag.2021.106002

Wakholi, C., Kandpal, L. M., Lee, H., Bae, H., Park, E., Kim, M. S., et al. (2018). Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics. *Sens Actuators B Chem.* 255, 498–507. doi: 10.1016/j.snb.2017.08.036

Wang, E., and Liu, M. (1996). A comparative biology study on the two medicinal astragalus. *Bull. Bot. Res.* 16, 87–93.

Wang, Z., Chen, J., Fan, Y., Cheng, Y., Wu, X., Zhang, J., et al. (2020). Evaluating photosynthetic pigment contents of maize using UVE-PLS based on continuous wavelet transform. *Comput. Electron Agric.* 169, 105160. doi: 10.1016/j.compag.2019.105160

Wang, H., Song, C., Sha, M., Liu, J., Li, L., and Zhang, Z. (2018). Discrimination of medicine radix astragali from different geographic origins using multiple spectroscopies combined with data fusion methods. *J. Appl. Spectrosc* 85, 313–319. doi: 10.1007/s10812-018-0650-4

Wang, J., Sun, L., Feng, G., Bai, H., Yang, J., Gai, Z., et al. (2022). Intelligent detection of hard seeds of snap bean based on hyperspectral imaging. *Spectrochim Acta A Mol. Biomol Spectrosc* 275, 121169. doi: 10.1016/j.saa.2022.121169

Wang, J., Zhang, H., and Jin, X. (2005). Morphologic characteristics and identification methods of seeds in astragalus membranaceus var. mongholicus and a. membranaceus. *Chin. Trad. Herb. Drugs* 36, 117–120.

Wang, L., Sun, D., Pu, H., and Zhu, Z. (2016). Application of hyperspectral imaging to discriminate the variety of maize seeds. *Food Anal. Methods* 9, 225–234. doi: 10.1007/s12161-015-0160-4

Wu, Y., Peng, S., Xie, Q., Han, Q., Zhang, G., and Sun, H. (2019). An improved weighted multiplicative scatter correction algorithm with the use of variable selection: Application to near-infrared spectra. *Chemometr. Intell. Lab. Syst.* 185, 114–121. doi: 10.1016/j.chemolab.2019.01.005

Xiao, Q., Bai, X., Gao, P., and He, Y. (2020). Application of convolutional neural network-based feature extraction and data fusion for geographical origin identification of radix astragali by visible/short-wave near-infrared and near infrared hyperspectral imaging. *Sensors (Switzerland)* 20, 1–14. doi: 10.3390/s20174940

Xu, Y., Tu, K., Cheng, Y., Hou, H., Cao, H., Dong, X., et al. (2021). Application of digital image analysis to the prediction of chlorophyll content in astragalus seeds. *Appl. Sci. (Switzerland)* 11, 8744. doi: 10.3390/app11188744

Yang, X., Hong, H., You, Z., and Cheng, F. (2015a). Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors (Switzerland)* 15, 15578–15594. doi: 10.3390/s150715578

Yang, Y., Sun, D., Pu, H., Wang, N., and Zhu, Z. (2015b). Rapid detection of anthocyanin content in lychee pericarp during storage using hyperspectral imaging coupled with model fusion. *Postharvest Biol. Technol.* 103, 55–65. doi: 10.1016/j.postharvbio.2015.02.008

Yang, J., Sun, L., Xing, W., Feng, G., Bai, H., and Wang, J. (2021a). Hyperspectral prediction of sugarbeet seed germination based on gauss kernel SVM. *Spectrochim Acta A Mol. Biomol Spectrosc* 253, 119585. doi: 10.1016/j.saa.2021.119585

Yang, R., Tian, H., and Kan, J. (2018). Classification of sugar beets based on hyperspectral and extreme learning machine methods. *Appl. Eng. Agric.* 34, 891–897. doi: 10.13031/aea.12903

Yang, Z., Yang, Y., and E, X., (2020). Research progress on the difference between astragalus mongholicus and astragalus membranaceus. *J. Chin. Med. Mater.* 43, 1261–1265. doi: 10.13863/j.issn1001-4454.2020.05.044

Yang, J., Yin, C., Miao, X., Meng, X., Liu, Z., and Hu, L. (2021b). Rapid discrimination of adulteration in *radix astragali* combining diffuse reflectance mid-infrared fourier transform spectroscopy with chemometrics. *Spectrochim Acta A Mol. Biomol Spectrosc* 248, 119251. doi: 10.1016/j.saa.2020.119251

Yang, S., Zhu, Q., Huang, M., and Qin, J. (2017). Hyperspectral image-based variety discrimination of maize seeds by using a multi-model strategy coupled with unsupervised joint skewness-based wavelength selection algorithm. *Food Anal. Methods* 10, 424–433. doi: 10.1007/s12161-016-0597-0

Yan, C., Nie, F., Dong, L., and Li, L. (2005). Identification of two astragalus seeds by scanning electron microscope and ultraviolet spectrum. *LISHIZHEN Med. Materia Med. Res.* 16, 720–721.

Yan, L., Wan, T., Zhang, Z., Wang, X., and Sun, W. (2001). Analysis on botanical characters of astragalus membranacens bunge. and a. membranacens var. mongolicus (Bunge) hsiao. *J. Inner Mongolia Agric. Univ.* 22, 71–77.

Zhang, T., Fan, S., Xiang, Y., Zhang, S., Wang, J., and Sun, Q. (2020). Non-destructive analysis of germination percentage, germination energy and simple vigour index on wheat seeds during storage by Vis/NIR and SWIR hyperspectral imaging. *Spectrochim Acta A Mol. Biomol Spectrosc* 239, 118488. doi: 10.1016/j.saa.2020.118488

Zhang, J., Feng, X., Liu, X., and He, Y. (2018a). Identification of hybrid okra seeds based on near-infrared hyperspectral imaging technology. *Appl. Sci. (Switzerland)* 8, 1793. doi: 10.3390/app8101793

Zhang, H., Ge, Y., Xie, X., Atefi, A., Wijewardane, N. K., and Thapa, S. (2022a). High throughput analysis of leaf chlorophyll content in sorghum using RGB, hyperspectral, and fluorescence imaging and sensor fusion. *Plant Methods* 18, 60. doi: 10.1186/s13007-022-00892-0

Zhang, H., Hou, Q., Luo, B., Tu, K., Zhao, C., and Sun, Q. (2022b). Detection of seed purity of hybrid wheat using reflectance and transmittance hyperspectral imaging technology. *Front. Plant Sci.* 13, 1015891. doi: 10.3389/fpls.2022.1015891

Zhang, R., Li, C., Zhang, M., and Rodgers, J. (2016). Shortwave infrared hyperspectral reflectance imaging for cotton foreign matter classification. *Comput. Electron Agric.* 127, 260–270. doi: 10.1016/j.compag.2016.06.023

Zhang, T., Wei, W., Zhao, B., Wang, R., Li, M., Yang, L., et al. (2018b). A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds. *Sensors (Switzerland)* 18, 813. doi: 10.3390/s18030813

Zhang, C., Yang, X., Wei, J., Chen, N., Xu, J., Bi, Y., et al. (2021). Ethnopharmacology, phytochemistry, pharmacology, toxicology and clinical applications of radix astragali. *Chin. J. Integr. Med.* 27, 229–240. doi: 10.1007/s11655-019-3032-8

Zhang, H., Zhang, L., Zhou, X., and Chen, Z. (2009). Comparative study on growth and development of *astragalus mongholicus* and *astragalus membranaceus*. *China J. Chin. Materia Med.* 34, 1591–1592.

Zhao, Y., Zhang, C., Zhu, S., Gao, P., Feng, L., and He, Y. (2018). Non-destructive and rapid variety discrimination and visualization of single grape seed using near-infrared hyperspectral imaging technique and multivariate analysis. *Molecules* 23, 1352. doi: 10.3390/molecules23061352

Zheng, S., Shang, X., Zeng, Y., and Wang, J. (2019). Research on identification of two plant origin of *astragali radix* based on specific molecular marker technology. *Modern Chin. Med.* 21, 307–311. doi: 10.13313/j.issn.1673-4890.20181117001

Zhou, Q., Huang, W., Fan, S., Zhao, F., Liang, D., and Tian, X. (2020a). Non-destructive discrimination of the variety of sweet maize seeds based on hyperspectral image coupled with wavelength selection algorithm. *Infrared Phys. Technol.* 109, 103418. doi: 10.1016/j.infrared.2020.103418

Zhou, S., Sun, L., Xing, W., Feng, G., Ji, Y., Yang, J., et al. (2020b). Hyperspectral imaging of beet seed germination prediction. *Infrared Phys. Technol.* 108, 103363. doi: 10.1016/j.infrared.2020.103363

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fpls.2022.1024360

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Zhaoyu Zhai,
Nanjing Agricultural University, China
Yecheng Lyu,
Volvo Car Technology USA, United States
Yanbo Huang,
United States Department of Agriculture (USDA), United States
Xiangjun Zou,
South China Agricultural University, China

*CORRESPONDENCE
Liqing Chen
lqchen@ahau.edu.cn

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 21 August 2022
ACCEPTED 31 October 2022
PUBLISHED 01 December 2022

CITATION
Li Z, Xie D, Liu L, Wang H and Chen L (2022) Inter-row information recognition of maize in the middle and late stages *via* LiDAR supplementary vision. *Front. Plant Sci.* 13:1024360. doi: 10.3389/fpls.2022.1024360

COPYRIGHT
© 2022 Li, Xie, Liu, Wang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Inter-row information recognition of maize in the middle and late stages *via* LiDAR supplementary vision

Zhiqiang Li[1,2], Dongbo Xie[1,2], Lichao Liu[1,2], Hai Wang[1,3] and Liqing Chen[1,2]*

[1]College of Engineering, Anhui Agricultural University, Hefei, China, [2]Anhui Intelligent Agricultural Machinery Equipment Engineering Laboratory, Hefei, China, [3]Discipline of Engineering and Energy, Murdoch University, Perth, WA, Australia

In the middle and late stages of maize, light is limited and non-maize obstacles exist. When a plant protection robot uses the traditional visual navigation method to obtain navigation information, some information will be missing. Therefore, this paper proposed a method using LiDAR (laser imaging, detection and ranging) point cloud data to supplement machine vision data for recognizing inter-row information in the middle and late stages of maize. Firstly, we improved the YOLOv5 (You Only Look Once, version 5) algorithm based on the characteristics of the actual maize inter-row environment in the middle and late stages by introducing MobileNetv2 and ECANet. Compared with that of YOLOv5, the frame rate of the improved YOLOv5 (Im-YOLOv5) increased by 17.91% and the weight size decreased by 55.56% when the average accuracy was reduced by only 0.35%, improving the detection performance and shortening the time of model reasoning. Secondly, we identified obstacles (such as stones and clods) between the rows using the LiDAR point cloud data to obtain auxiliary navigation information. Thirdly, the auxiliary navigation information was used to supplement the visual information, so that not only the recognition accuracy of the inter-row navigation information in the middle and late stages of maize was improved but also the basis of the stable and efficient operation of the inter-row plant protection robot was provided for these stages. The experimental results from a data acquisition robot equipped with a camera and a LiDAR sensor are presented to show the efficacy and remarkable performance of the proposed method.

KEYWORDS

inter-row information recognition, point cloud, maize plant protection, lidar, machine vision

1 Introduction

Maize is one of the five most productive cereals in the world (the other four being rice, wheat, soybean, and barley) (Patricio and Rieder, 2018) that is an important source of food crops and feed. In recent years, with the rapid increase in maize consumption, an efficient and intelligent maize production process has been required to increase productivity (Tang et al., 2018; Yang et al., 2022a). Inter-row navigation is a key to realizing the intelligence of maize planting. Pest control in the middle and late stages of maize determines the crop yield and quality. A small autonomous navigation plant protection robot is a good solution for plant protection in the middle and late stages of maize development (Li et al., 2019). However, in these stages, the high plant height (Chen et al., 2018), insufficient light, and several non-maize obstacles lead to a typical high-occlusion environment (Hiremath et al., 2014; Yang et al., 2022b). Commonly used navigation systems such as GPS (Global Positioning System) and BDS (BeiDou Navigation Satellite System) have shown poor signal quality in a high-occlusion environment (Gai et al., 2021); therefore, accurately obtaining navigation information between rows in the middle and late stages of maize has become the key issue to realizing the autonomous navigation of plant protection robots. At present, machine vision is the mainstream navigation method used to obtain inter-row navigation information in a high-occlusion environment (Radcliffe et al., 2018); that is, the RGB (red, green, and blue) camera acquires images of the maize stems, identifies maize stems through a trained model, and obtains position information so as to plan the navigation path. The convolutional neural network was used to train the robot to recognize the characteristics of maize stalks at the early growth stage, which was implemented on an inter-row information collection robot based on machine vision (Gu et al., 2020). Tang et al. reported the application and research progress of harvesting robots and vision technology in fruit picking (Tang et al., 2020). The authorsMachine vision technology was applied for the multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point (Wu et al., 2021); in addition, the improved YOLOv4 (You Only Look Once, version 4) micromodel and binocular stereo vision technology were applied for fruit detection and location (Wang et al., 2022; Tang et al., 2023). Zhang et al. proposed an inter-row information recognition algorithm for an intelligent agricultural robot based on binocular vision, where the effective inter-row navigation information was extracted by fusing the edge contour and height information of crop rows in the image (Zhang et al., 2020). By setting the region of interest, Yang et al. used machine vision to accurately identify the crop lines between rows in the early growth stage of maize and extracted the navigation path of the plant protection robot in real time (Yang et al., 2022a). However, the inter-row environment in the middle and late stages of maize is a typical high-occlusion environment, with higher plant height and dense branches and leaves, seriously blocking light (Liu et al., 2016; Xie et al., 2019). When the ambient light intensity is weak, information loss will occur when using machine vision to obtain inter-row navigation information (Chen et al., 2011). However, considering the fact that machine vision usually takes a certain feature of maize as the basis for the acquisition of information, recognizing multiple features at the same time will greatly reduce the recognition speed and also reduce the real-time performance of agricultural robots, taking non-maize obstacles into consideration (such as soil, bricks, and branches) in the middle and late stages of maize; it is, therefore, quite difficult to obtain all the inter-row information by using only a single feature.

Since LiDAR (laser imaging, detection and ranging) can obtain accurate point cloud data of objects according to the echo detection principle (Reiser et al., 2018; Wang et al., 2018; Jafari Malekabadi et al., 2019) and is less affected by light (Wang et al., 2022a; Wang et al., 2022b), it can supplement the missing information caused by the use of machine vision (Jeong et al., 2018; Aguiar et al., 2021). In order to solve the issue of information loss when a vision sensor was used to obtain information, a method using LiDAR supplement vision was proposed (Bae et al., 2021), which pooled the strength of each sensor and made up for the shortcomings of using a single sensor. Through the complementary process between vision and LiDAR (Morales et al., 2021; Mutz et al., 2021), the performance of adaptive cruise control was significantly improved; thus, a complementary method combining vision and LiDAR was developed in order to further improve the accuracy of unmanned aerial vehicle (UAV) navigation (Yu et al., 2021). Liu et al. proposed a new structure of LiDAR supplement vision in an end-to-end semantic segmentation network, which can effectively improve the performance of automatic driving (Liu et al., 2020). The above methods had good application effects in the field of autonomous driving (Chen et al., 2021; Yang et al., 2021; Zhang et al., 2021). Based on the above research, we believe that LiDAR supplement vision is an interesting and effective method to obtaining inter-row information in the middle and late stages of maize development.

Therefore, this paper proposed a method of using LiDAR point cloud data to supplement machine vision data for obtaining inter-row information in the middle and late stages of maize. We took the location of maize plants as the main navigation information and proposed an improved YOLOv5 (Im-YOLOv5) algorithm (Jubayer et al., 2021, p. 5) to identify maize plants and obtain the main navigation information. At the same time, we took the locations of stones, clods, and other obstacles as auxiliary navigation information, which were obtained through LiDAR. By the supplementary function of vision and LiDAR, the accuracy of the inter-row navigation information acquisition in the middle and late stages of maize can be improved. The proposed method provides a new and effective way to obtaining navigation

information between rows in the middle and late stages of maize under the condition of equal height occlusion.

The contributions of this article are summarized as follows:

1. A method of inter-row information recognition with a LiDAR supplement camera is proposed.
2. An Im-YOLOv5 model with efficient channel attention (ECA) and lightweight backbone network is established.
3. Auxiliary navigation information acquisition using LiDAR can reduce the loss of information.
4. The proposed method was tested and analyzed using a data acquisition robot.

2 Methods and materials

2.1 Composition of the test platform

The experimental platform and data acquisition system are shown in Figure 1. A personal computer (PC) was used as the upper computer to collect LiDAR and camera signals. The LiDAR model is VLP-16, the scanning distance was 100 m, the horizontal scanning angle was 270°, and the vertical scanning angle was ±15°. The camera model is NPX-GS650, the resolving power was 640*480, and the frame rate was 790.

2.2 Commercialization feasibility analysis

The data acquisition platform used in the test costs 490 RMB. The plant protection operation can be carried out by installing a pesticide applicator in the later stage, with the cost of the pesticide applicator about 100 RMB. The cost of the camera sensor was about 100 RMB, and that of the LiDAR sensor was about 5,000 RMB. Consequently, the cost of VLP-16 LiDAR represented a key issue affecting the commercialization of this recognition system. Therefore, our recognition system was applied to small autonomous navigation plant protection robots. The relatively low-cost of small plant protection robots, even with the application of this relatively high-precision recognition system, had a price advantage over UAVs.

3 Joint calibration of camera and LiDAR

In this paper, a monocular camera and VLP-16 LiDAR were used as the information fusion sensors. When the monocular camera and the LiDAR detect the same target, despite the range and angle information being the same, the detection results of the two sensors belong to different coordinate systems (Chen et al., 2021a). Therefore, in order to effectively realize the information supplementation of LiDAR to the camera, the coordinate system must be unified; that is, the detection results of the two sensors should be input into the same coordinate system and the relative pose between them should be calibrated at the same time so as to realize the data matching and correspondence between these two sensors.

It should be noted that the main task of the monocular camera calibration was to solve its extrinsic parameter matrix and intrinsic parameters. In this paper, the chessboard calibration method was used (Xu et al., 2022), with the



FIGURE 1
Data acquisition robot. *PC*, personal computer.

chessboard size being 400 mm × 550 mm and the grid size being 50 mm × 50; mm. We randomly took 21 chessboard pictures of different positions. The camera calibration error was less than 0.35 pixels and the overall mean error was 0.19 pixels, which means, according to reference, that the error met the calibration accuracy and that the calibration result has practical value (Xu et al., 2021). The internal parameters of the camera were as follows: focal length (f) = 25 mm, radial distortion parameter (k_1) = 0.012 mm, radial distortion parameter (k_2) = 0.009 mm, tangential distortion parameter (p_1) = −0.0838 mm, tangential distortion parameter (p_2) = 0.1514 mm, image center (u_0) = 972 mm, image center (v_0) = 1,296 mm, normalized focal length ($f_x = f/dx$) = 1,350.3 mm, and normalized focal length ($f_y = f/dy$) = 2,700.8 mm. On the basis of camera calibration, we carried out the joint calibration of the camera and LiDAR. The calibration principle is shown in Figure 2A. By matching the corner information of the chessboard picture taken by the camera to the corner information of the chessboard point cloud data obtained by LiDAR, a rigid transformation matrix from the point cloud data to the image can be obtained. During calibration, the camera and LiDAR were fixed on the data acquisition robot platform developed by the research group. After the joint calibration, the relative positions of the camera and LiDAR were saved and fixed. The calibration error is shown in Figure 2B. As indicated in Aguiar et al. (2021), the calibration error met the calibration accuracy, and the calibration result showed practical value. Through joint calibration, the rigid transformation matrix of the point cloud projection to the image is obtained from Equations (1) and (2).

$$R_{\text{lidar}} = \begin{bmatrix} 0.9998 & 0.0176 & 0.0047 \\ -0.0032 & -0.0807 & -0.9967 \\ 0.0179 & -0.9966 & 0.0806 \end{bmatrix} \quad (1)$$

$$T_{\text{lidar}} = \begin{bmatrix} 0.0468 \\ 0.1139 \\ -0.2667 \end{bmatrix} \quad (2)$$

4 Navigation information acquisition based on LiDAR supplement vision

As mentioned in Section 1, machine vision usually takes a single feature of the plant as the basis of recognition. In this paper, the maize stem about 10 cm above the ground surface was taken as the machine vision recognition feature. It should be noted that taking the maize stem as the identification feature will cause lack of information on the other non-maize obstacles (such as stones and clods). In order to solve the issue of missing information when using machine vision to acquire navigation information, this paper proposed a method of inter-row navigation information acquisition in the middle and late stages of maize based on LiDAR supplement vision. The detailed principle is shown in Figure 3. The machine vision datasets were trained using the Im-YOLOv5 algorithm to identify the stem of the maize and, subsequently, to obtain the main navigation information. The point cloud data of the inter-row environment in the middle and late stages of maize were obtained using LiDAR to gather auxiliary navigation information. It should be noted that the method proposed in this paper obtained inter-line information through LiDAR-assisted cameras; therefore, spatial data fusion was used. After establishing the precise coordinate conversion relationship among the radar coordinate systems—a three-dimensional world coordinate system—a camera coordinate system, an image coordinate system, and a pixel coordinate system—the spatial position information of the obstacles in the point cloud data can be matched to the visual image.

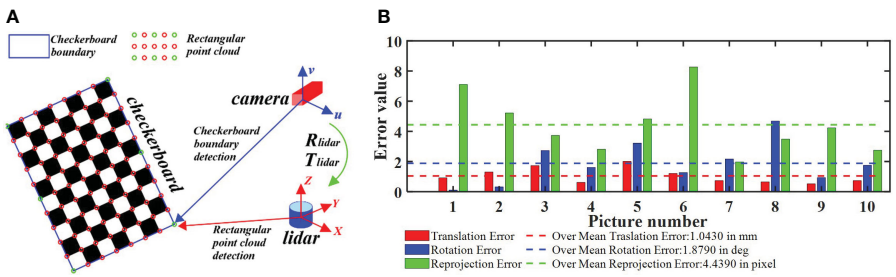


FIGURE 2
Camera–LiDAR (laser imaging, detection and ranging) joint calibration process. **(A)** Principle of joint calibration. **(B)** Joint calibration error. By matching the corner information of the chessboard picture taken by the camera to the corner information of the chessboard point cloud data obtained by LiDAR, the rigid transformation matrix from the point cloud data to the image can be obtained.

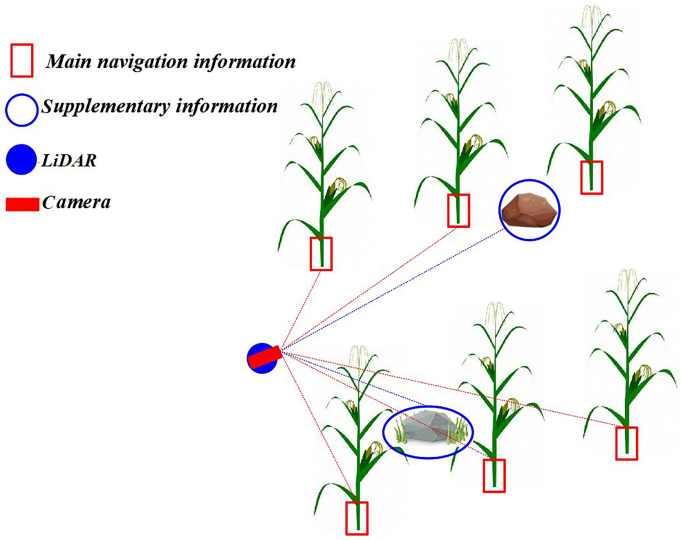


FIGURE 3
Principle of navigation information acquisition based on LiDAR (laser imaging, detection, and ranging) supplement camera. The machine vision datasets were trained using the improved YOLOv5 (Im-YOLOv5) algorithm to identify the stem of the maize and then obtain the main navigation information, while LiDAR was used to obtain auxiliary navigation information.

4.1 Main navigation information acquisition with the improved YOLOv5

YOLO models have a real-time detection speed, but require a powerful GPU (graphic processing unit) and a large amount of memory when training, limiting their use on most computers. The large size of the model after training can also increase the hardware requirements on mobile devices. Ideally, a detection model would meet the requirements of detection accuracy and real-time detection speed of maize stems, without high hardware requirements. The YOLOv5 model is a lightweight version of YOLO, has fewer layers and faster detection speed, can be used on portable devices, and requires fewer GPU resources for training (Tang et al., 2023). Therefore, the goal of this work was to build on the YOLOv5 model and apply the improved model for the detection of maize stems. The main idea for improving YOLOv5 was to lighten its backbone network through MobileNetv2 and introduce the ECANet attention mechanism to improve the recognition accuracy and robustness of the model.

4.1.1 Lightweight Backbone network

This paper used MobileNetv2 (Zhou et al., 2020) to replace the backbone network of YOLOv5 for the extraction of maize stem images with effective characteristics. In order to enhance the adaptability of the network to the task of recognizing maize stem features and fully extract features, a progressive classifier was designed in this paper to enhance the network's recognition ability of the corn rhizome. The original MobileNetV2 network was primarily used to deal with more than 1,000 types of targets on the ImageNet dataset, while this paper only targeted maize stems. Therefore, in order to better extract the characteristics of maize stems and improve the recognition ability of the network on maize stems, we the classifier of the network was redesigned, which included two convolution layers, one global pooling layer, and one output layer (convolution layer).

The main task of the classifier was to efficiently convert the extracted maize stem features into specific classification results. As shown in Figure 4, two convolution kernels with different scales were selected to replace a single convolution kernel in the original classifier in order to perform the compression and conversion operations of the feature map. The size of the first convolution kernel was 1×1. It was mainly responsible for the channel number compression of the feature map. In order to avoid the loss of a large number of useful features caused by a large compression ratio, the second convolution was used mainly for the size compression of the feature map to avoid fluctuations in the subsequent global pooling on a large feature map. Comparison of the Im-YOLOv5 network based on MobileNetv2 with the original YOLOv5 network showed that the model parameters decreased from 64,040,001 to 39,062,013 and the parameters decreased by 39%.

At the same time, Im-YOLOv5 used CIOU_Loss [complete intersection over union (IOU) loss] to replace GIOU_Loss (generalized IOU loss) as the loss function of the bounding box and used binary cross-entropy and logits loss function to calculate the loss of class probability and target score, defined as follows.

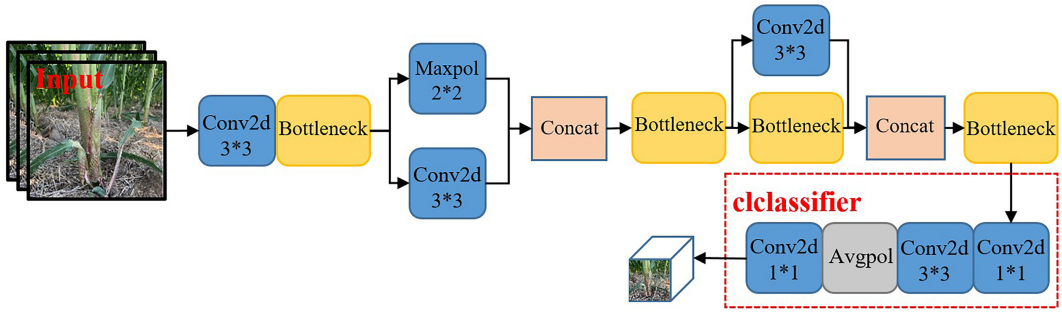


FIGURE 4
MobileNetv2 network structure.

$$\mathbf{GIOU} = \mathbf{IOU} - \left|\frac{C - (A \cup B)}{|C||}\right. \tag{3}$$

$$\mathbf{IOU} = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

$$\mathbf{CIOU} = \mathbf{IOU} - \frac{\rho^2(b, b^{gt})}{c^2} - av \tag{5}$$

In Equations (3) and (4), A and B are the prediction box and the real box, respectively; IOU is the intersection ratio of the prediction box and the real box; and C is the minimum circumscribed rectangle of the prediction box and the target box. However, Equations (3) and (4), considering only the overlap rate between the prediction box and the target box, cannot describe well the regression problem of the target box. When the prediction box is inside the target box and the size of the prediction box is the same, GIOU will degenerate into IOU, which cannot distinguish the corresponding positions of the prediction box in each target box, resulting in error detection and leak detection. Equation (5) is the calculation formula of CIOU, where $a = v/(1\text{-IOU})v$ is an equilibrium parameter that does not participate in gradient calculation; v = 4/π^2(arctan (W^{gt}/H^{gt}) – arctan $(W/H))^2$ is a parameter used to measure the consistency of the length-width ratio; b is the forecast box; b^{gt} is the realistic box; ρ is the Euclidean distance; and c is the diagonal length of the minimum bounding box. It can be seen from Equation (5) that the CIOU comprehensively considers the overlapping area, center point distance, aspect ratio, and other factors of the target and prediction boxes and solves the shortcoming of the GIOU loss function, making the regression process of the target box more stable, with faster convergence speed and higher convergence accuracy.

4.1.2 Introducing the attention mechanism

In order to improve the recognition accuracy and robustness of the algorithm in the case of a large number of maize stems and mutual occlusion between stems, efficient channel attention (ECA) was introduced (Xue et al., 2022). It should be noted that, although the introduction of ECANet into convolutional neural networks has shown better performance improvements, ECANet only considers the local dependence between the current channel of the feature map and several adjacent channels, which inevitably loses the global dependence between the current channel and other long-distance channels. On the basis of ECANet, we added a new branch (shown in the dashed box in Figure 5) that has undergone channel-level global average pooling and is disrupted. This branch randomly rearranges the channel order of the feature map after undergoing channel-level global average pooling, so the long-distance channel before disruption may become its adjacent channel. After obtaining the local dependencies between the current channel of the new feature map and its new k adjacent channels, weighting the two branches can obtain more interaction information between channels.

In this paper, suppose that the feature vector of the input feature after convolution is $x \in R^{W\times H\times C}$, where W, H, and C respectively represent the width, height, and channel size of the feature vector. The global average pooling of the channel dimension can be expressed as:

$$y = g(x) = \frac{1}{WH}\sum_{i=1,j=1}^{W,H} x_{ij} \tag{6}$$

Then, in ECANet, the feature vector inputs by the two branches can be expressed as:

$$ys = S(g(x)) = S\left(\frac{1}{WH}\sum_{i=1,j=1}^{W,H} x_{ij}\right) \tag{7}$$

$$yg = g(x) = \frac{1}{WH}\sum_{i=1,j=1}^{W,H} x_{ij} \tag{8}$$

where ys represents the vector obtained after global average pooling and disrupting the sequential branching of channels; yg

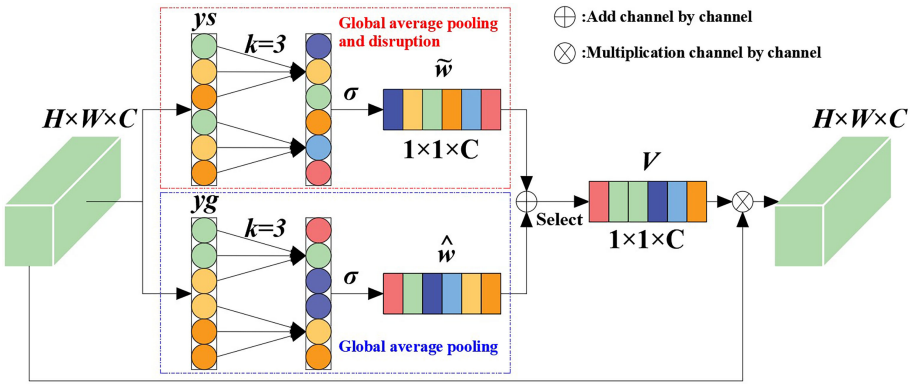


FIGURE 5
ECANet channel attention.

represents the vector obtained after global average pooling and branching; and S is a channel-disrupting operation. Given that the feature vector without dimension reduction is $y \in R^C$, the inter-channel weight calculation using the channel attention module can be expressed as:

$$\omega = \sigma(W_k y) \tag{9}$$

where $\sigma(x) = 1/(1+e^{-x})$ is the sigmoid activation function and W_k is the parameter matrix for calculating channel attention using ECANet.

We took MobileNetv2 (Zhou et al., 2020) as the backbone model, combined YOLOv5 with the SeNet and ECANet modules (Hassanin et al., 2022), and carried out maize stem recognition experiments. The test results are shown in Table 1. ECANet showed better performance compared toSeNet, indicating that ECANet can improve the performance of YOLOv5 with less computational costs. At the same time, ECANet was more competitive than SeNet, and the model complexity was also lower.

In this work, the ECANet attention mechanism was first placed on the enhanced feature extraction network and the attention mechanism added on the three effective feature layers extracted from the backbone network. Regarding the problems of information attenuation, the aliasing effect of cross-scale fusion and the inherent defects of channel reduction in the feature pyramid network (FPN) in YOLOv5, in this paper, we added the ECANet attention mechanism to the sampling results on FPN in order to reduce information loss and optimize the integration characteristics on each layer. By introducing the ECANet attention mechanism, Im-YOLOv5 can better fit the relevant feature information between the target channels, ignore and suppress useless information, and make the model focus more on training the specific category of maize stems, strengthening it and improving its detection performance. The specific structure of the Im-YOLOv5 algorithm is shown in Figure 6.

4.2 Auxiliary navigation information acquisition by LiDAR

Because of the obvious color and structural characteristics of maize stems, we trained the Im-YOLOv5 model to only detect maize stems when the main navigation information was obtained through machine vision. However, the actual non-maize obstacles were mainly soil blocks and stones, and the color and shape characteristics of such obstacles are relatively close to the ground color, which greatly increased the difficulty of Im-YOLOv5 model training. At the same time, recognizing multiple features simultaneously by machine vision will also reduce the recognition speed to a certain extent. Under this condition, it is necessary to obtain point cloud information using LiDAR to supplement machine vision.

TABLE 1 Comparison of the recognition performance (in percent) of the YOLOv5 model integrating different attention modules.

Method	*P*	*R* (%)	FPS (%)	F_1 (%)	mAP (%)
MobileNetv2	89.5	94.1	34.2	93.7	91.37
MobileNetv2+SeNet	93.2	91.4	63.7	91.2	97.25
MobileNetv2+ECANet	96.7	82.3	79.6	86.3	96.98

P, comparison of accuracy; R, recall; F_1, harmonic average; FPS, frame rate; mAP, mean average precision.

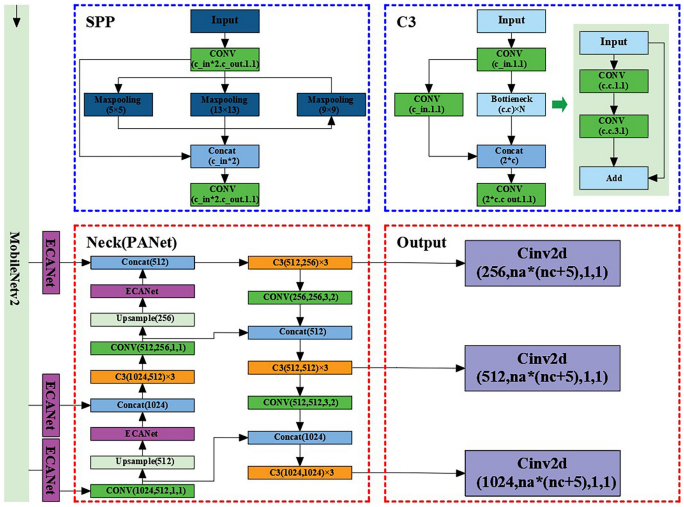


FIGURE 6
Improved YOLOv5 (Im-YOLOv5) architecture.

4.2.1 Determination of the effective point cloud range

Since the camera and LiDAR were fixed on the data acquisition robot platform, when the robot is walking between lines during data acquisition, it is necessary to determine the effective data range of the LiDAR point cloud according to the shooting angle range of the camera, as shown in Figure 7A.

Note that, in Figure 7A, θ_e is the camera shooting angle range, θ_e is the scanning angle of LiDAR, and d is the width of the robot. Therefore, the range of the effective point cloud data collected by LiDAR is the sector area, where r is the radius of the sector with the angle of θ_e and is defined as:

$$r = \frac{d}{\cos\left(\frac{\pi}{2} - \frac{\theta_i}{2} - \theta_e\right)} \tag{10}$$

4.2.2 Coordinate conversion of the auxiliary navigation information

Through the joint calibration of the camera and LiDAR in the above section, the camera external parameter matrix (R, T), the camera internal parameter, and the rigid conversion matrix (R_{lidar}, T_{lidar}), of the camera and LiDAR sensor information were obtained.

In order to supplement the main navigation information with the auxiliary navigation information, it is essential to

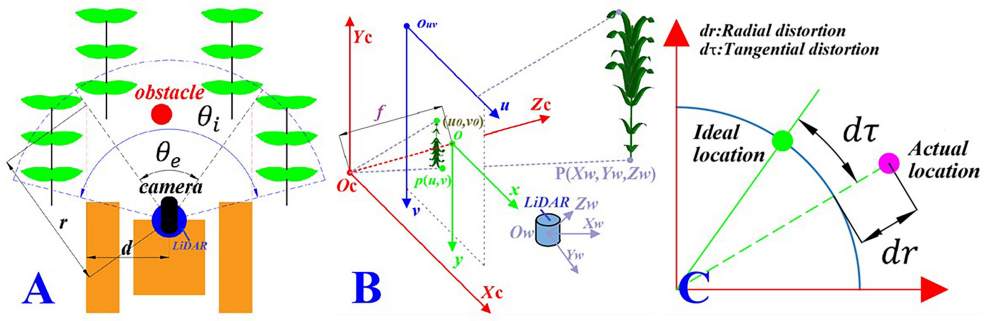


FIGURE 7
Camera–LiDAR (laser imaging, detection, and ranging) joint calibration process. **(A)** Effective data range. θ_e is the camera shooting angle range, θ_i is the scanning angle of LiDAR, and the overlapping area is the effective point cloud range. **(B)** Coordinate transformation. O_w - $X_wY_wZ_w$is the LiDAR coordinate system, O_c – $X_cY_cZ_c$is the camera coordinate system, o - $xy$is the image coordinate system, and O_{uv} – uv is the pixel coordinate system. **(C)** Distortion error. dr and dτ are the radial distortion and the tangential distortion of the camera, respectively.

establish a conversion model between sensors. Through the established transformation model, the points in the world coordinate system scanned by LiDAR were projected into the pixel coordinate system of the camera to realize the supplementation of the point cloud data to the visual information according to the pinhole camera model, as shown in Figure 7B. Note that, in Figure 7B, P is the point on the real object, p is the imaging point of P in the image, (x, y) are the coordinates of p in the image coordinate system, (u, v) are the coordinates of p in the pixel coordinate system, and f is the focal length of the camera, where $f = || o - 0_c||$ (in millimeters). The corresponding relationship between a point $P(X_w, Y_w, Z_w)$ in the real world obtained by LiDAR and the corresponding point $p(u, v)$ in the camera pixel coordinate system can be expressed as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{\mathbf{lidar}} & T_{\mathbf{lidar}} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_{\mathbf{w}} \\ Y_{\mathbf{w}} \\ Z_{\mathbf{w}} \\ 1 \end{bmatrix} \tag{11}$$

According to the principle of LiDAR scanning, the point cloud data obtained by LiDAR are in the form of polar coordinates. Therefore, the distance and angle information of the point cloud data under polar coordinates were converted into the three-dimensional coordinate point information under the LiDAR ontology coordinate system. The conversion formula was as follows:

$$\begin{cases} X_{\mathbf{w}} = \rho \cdot \mathbf{cos}\alpha \cdot \mathbf{cos}\theta \\ Z_{\mathbf{w}} = \rho \cdot \mathbf{cos}\alpha \cdot \mathbf{sin}\theta \\ Y_{\mathbf{w}} = \rho \cdot \mathbf{sin}\alpha \end{cases} \tag{12}$$

where ρ is the distance between the scanning point and the LiDAR;α is the elevation angle of the scanning line at the scanning point, namely, the angle in the vertical direction; and θ is the heading angle in the horizontal direction.

In order to eliminate the camera imaging distortion error caused by the larger deflection of light away from the lens center and the lens not being completely parallel to the image plane, as shown in Figure 7C, we corrected the distortion of Equation (11) with the correction formula, given as follows (Chen et al., 2021b):

Radial distortion correction:

$$\begin{cases} u' = u(1 + k_1 r^2 + k_2 r^4) \\ v' = v(1 + k_1 r^2 + k_2 r^4) \end{cases} \tag{13}$$

Tangential distortion correction:

$$\begin{cases} u'' = u' + (2p_1 v' + p_2(r^2 + 2u'^2) \\ v'' = v' + (2p_2 u' + p_1(r^2 + 2v'^2) \end{cases} \tag{14}$$

Where k_1 and k_2 are the radial correction parameters; p_1 and p_2 are the tangential correction parameters; $u''$and v' re the radially corrected pixel coordinates; and u'' and v'' are the tangentially corrected pixel coordinates.

The corresponding relationship between the point in the world coordinate system obtained by LiDAR and the camera pixel coordinate system is established through Equations (10)–(14). According to the established coordinate transformation model, the LiDAR point cloud data can be converted to the image space for the purpose of supplementation between machine vision and LiDAR.

4.2.3 Feature recognition of point cloud based on PointNet

Because of the irregular format of the point cloud, it is difficult to extract its feature, but with the proposal of the PointNet model (Jing et al., 2021), this problem was solved. In this paper, the features of the non-maize obstacles in the middle and late stages of maize were extracted through PointNet, and their location information taken as the output. Note that we also performed the following work before using the PointNet model for training. The principle is shown in Figure 8.

4.2.3.1 Ground segmentation

In order to obtain auxiliary navigation information from the LiDAR point cloud data, the ground point cloud must be

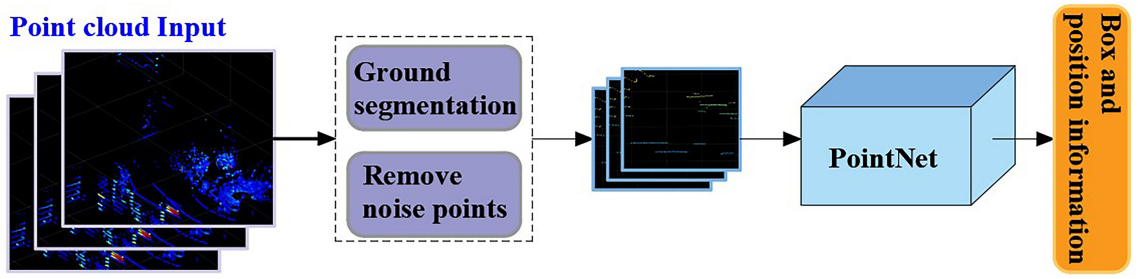


FIGURE 8
The principle of auxiliary navigation.

segmented first. In this work, the RANSAC (random sample consensus) algorithm was adopted to segment the collected point cloud data.

The unique plane can be determined by randomly selecting three non-collinear sample points (x_a, x_b, x_c) in the point cloud.

$$n_i \cdot x + d_i = 0 \tag{15}$$

$$n_i = (x_b - x_a)(x_c - x_a) \tag{16}$$

$$d_i = -n \cdot x_a \tag{17}$$

Where n_i is the normal vector of the plane model and d_i is the pitch of the plane model. Then, the distance from any sample point x_i in the point cloud to the plane model is given by

$$r_i = (n_i \cdot x_i + d_i)^2 \tag{18}$$

Let the distance threshold be T, when $r_i<T$. The sample point x_i is the internal point; otherwise, it is the external point. Let N be the number of internal points with

$$N = \mathbf{NUM}(x_i), r_i < T \tag{19}$$

sNote that Equations (15)–(19) show a calculation process, but N is not necessarily the maximum value at this time; hence, an iterative calculation is needed. Let the number of iterations be k_c. When N takes the maximum value, $N_{\max}$, in the iterative process, the plane model corresponding to n_{best} and d_{best} is the best-fitting ground.

4.2.3.2 Removing noise points caused by maize leaves

LiDAR was mainly used to identify obstacles other than maize leaves. In order to reduce the difficulty of model training, the point cloud data of maize leaves were deleted. This technology depends on the analysis of the z-coordinate distribution of each point cloud. In general, the height of obstacles such as soil blocks and stones is less than 10 cm. Therefore, when we trained the model sexually, we deleted the point cloud with a z-coordinate greater than 10 cm in the θ_e range.

5 Experiments and discussions

The focus of this paper was navigation information acquisition. Navigation information can be used for path planning to guide the robot to drive autonomously and can also be used as the basis for the adjustment of the driving state of the robot, such as reducing the driving speed when detecting rocks or large clods. We provided the results of the information acquisition experiment.

5.1 Main navigation information acquisition experiment

We verified the recognition performance of the Im-YOLOv5 for the main navigation information from two aspects: model training and detection results. In order to facilitate comparisons, we also provided the test results of YOLOv5 and Faster-RCNN (faster region-based convolutional network). The datasets used in the experiment were collected by the Anhui Intelligent Agricultural Machinery Equipment Engineering Laboratory. It should be noted that, in order for each model to perform best on the datasets, we adjusted the parameters of each model separately to select the appropriate hyperparameters. The initial hyperparameter settings of each algorithm are shown in Table 2. We divided the train set, test set, and verification set according to an 8:1:1 ratio, and the dataset contained 3,000 images.

The model training and validation loss rate curves are shown in Figure 9. From the figure, it can be seen that the loss rate tends to stabilize with the increase of iterations, finally converging to the fixed value; this indicates that the model has reached the optimal effect. The debugged model showed good fitting and generalization ability for the maize stem datasets. Note that, due to the Im-YOLOv5 having an improved loss function, the initial loss value of the model was about 0.38, which was the lowest among the three models, and the convergence speed was accelerated.

The P (comparison of accuracy), R (recall), F_1 (harmonic average), FPS (frame rate), and mAP (mean average precision)

TABLE 2 Target detection hyperparameter setting.

Parameter	Im-YOLOv5	YOLOv5	Faster-RCNN
Backbone network	MobileNetv2	Backbone	Resnet50
Training size	640 × 640	640 × 640	640 × 640
Batch size	16	16	8
No. of categories	5	5	5
Initial learning rate	1e−2	1e−2	1e−4
No. of iterations	100	100	100

Im-YOLOv5, improved You Only Look Once, version 5; Faster-RCNN, faster region-based convolutional network.

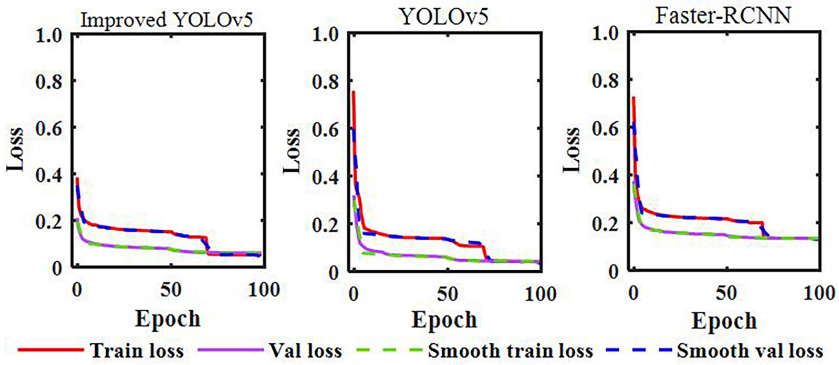


FIGURE 9
Model training and validation loss rate curves.

values for Im-YOLOv5, YOLOv5, and Faster-RCNN are shown in Table 3. From the table, it can be seen that Im-YOLOv5 had the highest accuracy rate, followed by YOLOv5; the accuracy rate of Faster-RCNN was low. With the lightweight backbone network, the FPS of Im-YOLOv5 was the highest, and the weight was greatly reduced. While meeting the real-time requirements, the detection speed of a single image was also the fastest and the detection performance was the best. Compared with that of YOLOv5, the FPS of Im-YOLOv5 was increased by 17.91% and the model size reduced by 55.56% when the mAP was reduced by only 0.35%, which improved the detection performance and shortened the model reasoning time. From the datasets, we selected a number of inter-row images of maize in the middle and late stages for testing, as shown in Figure 10. For the same image, Im-YOLOv5 was able to identify most maize stems, even those that were partially covered. At the same time, the detection confidence of Im-YOLOv5 and YOLOv5 was high, but that of Faster-RCNN was relatively low.

5.2 Auxiliary navigation information supplements the main navigation information experiment

In the experiments, the practical feasibility of the proposed inter-row navigation information acquisition method was verified based on LiDAR point cloud data-supplemented machine vision in the middle and late stages of maize. Considering the current coronavirus outbreak, conducting large-scale field experiments had been difficult. Therefore, an artificial maize plant model was used to set up the simulation test environment for verifying the feasibility of the designed method. Figure 11A shows the test environment using the maize plant model. Investigation of maize planting in Anhui Province revealed that the row spacing for maize plants is about 50–80 cm and that plant spacing is about 20–40 cm. Therefore, the row spacing in the maize plant model was set to 65 cm and the plant spacing to 25 cm. At the same time, a number of non-maize obstacles were also set in the experiments. For the purpose of data acquisition in this work, the data acquisition robot was developed by Anhui Intelligent Agricultural Machinery and Equipment Engineering Laboratory at Anhui Agricultural University.

During the experiments, the required main navigation information was the position information of maize plants, while the required auxiliary navigation information was the position information of the non-maize obstacles. We set up six maize plant models and three non-maize obstacles and randomly set the locations of the obstacles. Subsequently, we conducted 10 information acquisition experiments at distances of 1,000, 2,000, and 3,000 mm from the data acquisition robot to the front row of the maize plant model. The test results are shown in Figures 11B, C.

TABLE 3 Model evaluation.

Model	*P* (%)	*R* (%)	F_1 (%)	FPS	Model size (M)	mAP (%)
Im-YOLOv5	97	81	88	78	12	96.12
YOLOv5	93	90	93	66	27	96.48
Faster-RCNN	76	92	82	20	108	90.52

P, comparison of accuracy; R, recall; F_1, harmonic average; FPS, frame rate; mAP, mean average precision.



FIGURE 10
Results of stem detection. **(A)** Improved You Only Look Once, version 5 algorithm (Im-YOLOv5). **(B)** YOLOv5. **(C)** Faster region-based convolutional network (Faster-RCNN).

5.3 Discussions

Generally, visual navigation between rows in the middle and late stages of maize extracts the maize characteristics and then fits the navigation path. If the camera was only used to obtain information based on the maize characteristics in the recognition stage, information on the non-maize obstacles between rows in the middle and late stages of maize is missed, as shown in Figures 11B, C. With the introduction of the Im-YOLOv5 stem recognition algorithm, sufficient training for

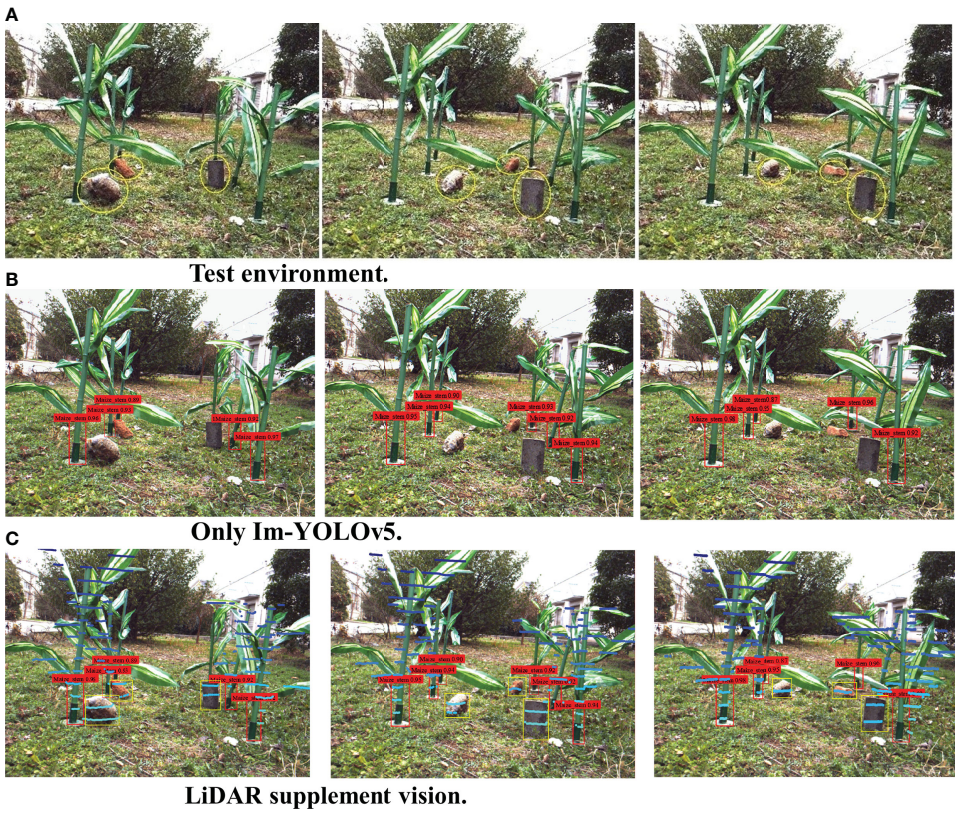


FIGURE 11
(A) Test environment. **(B)** Only the improved You Only Look Once, version 5 algorithm (Im-YOLOv5). **(C)** Laser imaging, detection, and ranging (LiDAR) supplement vision.

maize stem recognition has become exceptionally accurate; however, the non-maize obstacle recognition rate was almost zero only for Im-YOLOv5, which is extremely fatal for the actual operation safety of plant protection robots in the middle and late stages of maize.

When using LiDAR to obtain auxiliary navigation information in order to supplement the main navigation information obtained by machine vision, the issue of missing information can be properly solved, with the safety of the planned navigation path under this condition being greatly improved. However, due to the recognition accuracy of the 16-line LiDAR and the error of the camera–LiDAR joint calibration, the information recognition effect was not very satisfactory when the obstacle is far away and is too small. With increasing distance between the data acquisition robot and the maize plant, the number of maize plant models can be stably maintained, which means that the identification of the main navigation information is also stable. However, recognition of the number of non-maize obstacles showed a downward trend, indicating that the recognition accuracy using the auxiliary navigation information was reduced. In view of these issues, we will be using the 32-line or the 64-line LiDAR, both with higher accuracy, in future experiments.

6 Conclusion

In order to solve the problem of missing information when using machine vision for inter-row navigation in the middle and late stages of maize, this paper has proposed a method using LiDAR point cloud data to supplement machine vision in order to obtain more accurate inter-row information in the middle and late stages of maize. Through training of the machine vision datasets with the Im-YOLOv5 model, the main navigation information was obtained by identifying maize plants between the rows of maize in the middle and late stages. As a supplement to the main navigation information acquired by machine vision, LiDAR has been used to provide additional information to identify information on other non-crop obstacles as auxiliary navigation information. Not only was the accuracy of information recognition improved, but technical support for planning a safe navigation path can also be provided. Experimental results from the data acquisition robot equipped with a camera and a LiDAR sensor have demonstrated the validity and the good inter-row navigation recognition performance of the proposed method for the middle and late stages of maize. However, with the improvement in the accuracy of LiDAR, cost is the key problem restricting the commercialization of this recognition system. Therefore, we hope that our recognition system can be applied in small autonomous navigation plant protection robots, as the relatively low cost of small plant protection robots, even with the application of this relatively high-precision recognition system, has a price advantage over UAVs. The navigation information can be used for path planning to guide robots to drive autonomously and can also be used as the basis for the adjustment of the driving state of robots, such as in reducing the driving speed when detecting rocks or large clods. Therefore, in subsequent research, we will focus on path planning between maize rows and the control of the driving state of robots.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ZL: Software, visualization, investigation, and writing—original draft. DX and LL: Investigation. HW: Writing—review and editing. LC: Conceptualization, methodology, writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the National Natural Science Foundation of China under grant no. 52175212.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aguiar, A., Oliveira, M., Pedrosa, E. F., and Santos, F. (2021). A camera to LiDAR calibration approach through the optimization of atomic transformations. *Expert Syst. Appl.* 176, 114894. doi: 10.1016/j.eswa.2021.114894

Bae, H., Lee, G., Yang, J., Shin, G., Lim, Y., and Choi, G. (2021). Estimation of closest in-path vehicle (CIPV) by low-channel LiDAR and camera sensor fusion for autonomous vehicle. *Sensors*. 21, 3124. doi: 10.3390/s21093124

Chen, X., Chen, Y., Song, X., Liang, W., and Wang, Y. (2021a). Calibration of stereo cameras with a marked-crossed fringe pattern. *Opt. Lasers Eng.* 147, 106733. doi: 10.1016/j.optlaseng.2021.106733

Chen, K.-W., Lai, C.-C., Lee, P.-J., Chen, C.-S., and Hung, Y.-P. (2011). Adaptive learning for target tracking and true linking discovering across multiple non-overlapping cameras. *IEEE Trans. Multimedia* 13, 625–638. doi: 10.1109/TMM.2011.2131639

Chen, L., Li, Z., Yang, J., and Song, Y. (2021). Lateral stability control of four-Wheel-Drive electric vehicle based on coordinated control of torque distribution and ESP differential braking. *Actuators* 10, 135. doi: 10.3390/act10060135

Chen, X., Song, X., Cheng, L., Ran, Z., Wang, Y., Liang, W., et al. (2021b). Flexible calibration method of electronically focus-tunable lenses. *IEEE Trans. Instrum. Meas* 70, 5013210. doi: 10.1109/TIM.2021.3097412

Chen, L., Wang, P., Zhang, P., Zheng, Q., He, J., and Wang, Q. (2018). Performance analysis and test of a maize inter-row self-propelled thermal fogger chassis. *Int. J. Agric. Biol. Eng.* 11, 100–107. doi: 10.25165/j.ijabe.20181105.3607

Gai, J., Xiang, L., and Tang, L. (2021). Using a depth camera for crop row detection and mapping for under-canopy navigation of agricultural robotic vehicle. *Comput. Electron. Agric.* 188, 106301. doi: 10.1016/j.compag.2021.106301

Gu, Y., Li, Z., Zhang, Z., Li, J., and Chen, L. (2020). Path tracking control of field information-collecting robot based on improved convolutional neural network algorithm. *Sensors* 20, 797. doi: 10.3390/s20030797

Hassanin, M., Anwar, S., Radwan, I., Khan, F. S., and Mian, A. (2022). Visual attention methods in deep learning: An in-depth survey. *arXiv*. doi: 10.48550/arXiv.2204.07756

Hiremath, S. A., van der Heijden, G. W. A. M., van Evert, F. K., Stein, A., and ter Braak, C. J. F. (2014). Laser range finder model for autonomous navigation of a robot in a maize field using a particle filter. *Comput. Electron. Agric.* 100, 41–50. doi: 10.1016/j.compag.2013.10.005

Jafari Malekabadi, A., Khojastehpour, M., and Emadi, B. (2019). Disparity map computation of tree using stereo vision system and effects of canopy shapes and foliage density. *Comput. Electron. Agric.* 156, 627–644. doi: 10.1016/j.compag.2018.12.022

Jeong, J., Yoon, T. S., and Park, J. B. (2018). Multimodal sensor-based semantic 3D mapping for a Large-scale environment. *Expert Syst. Appl.* 105, 1–10. doi: 10.1016/j.eswa.2018.03.051

Jing, Z., Guan, H., Zhao, P., Li, D., Yu, Y., Zang, Y., et al. (2021). Multispectral LiDAR point cloud classification using SE-PointNet plus. *Remote Sens.* 13, 2516. doi: 10.3390/rs13132516

Jubayer, F., Soeb, J. A., Mojumder, A. N., Paul, M. K., Barua, P., Kayshar, S., et al. (2021). Detection of mold on the food surface using YOLOv5. *Curr. Res. Food Sci.* 4, 724–728. doi: 10.1016/j.crfs.2021.10.003

Li, Z., Chen, L., Zheng, Q., Dou, X., and Yang, L. (2019). Control of a path following caterpillar robot based on a sliding mode variable structure algorithm. *Biosyst. Eng.* 186, 293–306. doi: 10.1016/j.biosystemseng.2019.07.004

Liu, L., Mei, T., Niu, R., Wang, J., Liu, Y., and Chu, S. (2016). RBF-based monocular vision navigation for small vehicles in narrow space below maize canopy. *Appl. Sci.-Basel* 6, 182. doi: 10.3390/app6060182

Liu, H., Yao, Y., Sun, Z., Li, X., Jia, K., and Tang, Z. (2020). Road segmentation with image-LiDAR data fusion in deep neural network. *Multimed. Tools Appl.* 79, 35503–35518. doi: 10.1007/s11042-019-07870-0

Morales, J., Vázquez-Martín, R., Mandow, A., Morilla-Cabello, D., and García-Cerezo, A. (2021). The UMA-SAR dataset: Multimodal data collection from a ground vehicle during outdoor disaster response training exercises. *Int. J. Robotics Res.* 40, 27836492110049. doi: 10.1177/02783649211004959

Mutz, F. W., Oliveira-Santos, T., Forechi, A., Komati, K. S., Badue, C., Frana, F., et al. (2021). What is the best grid-map for self-driving cars localization? an evaluation under diverse types of illumination, traffic, and environment. *Expert Syst. Appl.* 179, 115077. doi: 10.1016/J.ESWA.2021.115077

Patricio, D. I., and Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* 153, 69–81. doi: 10.1016/j.compag.2018.08.001

Radcliffe, J., Cox, J., and Bulanon, D. M. (2018). Machine vision for orchard navigation. *Comput. Ind.* 98, 165–171. doi: 10.1016/j.compind.2018.03.008

Reiser, D., Vázquez-Arellano, M., Paraforos, D. S., Garrido-Izard, M., and Griepentrog, H. W. (2018). Iterative individual plant clustering in maize with assembled 2D LiDAR data. *Comput. Industry* 99, 42–52. doi: 10.1016/j.compind.2018.03.023

Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., et al. (2020). Recognition and localization methods for vision-based fruit picking robots: A review. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00510

Tang, D., Feng, Y., Gong, D., Hao, W., and Cui, N. (2018). Evaluation of artificial intelligence models for actual crop evapotranspiration modeling in mulched and non-mulched maize croplands. *Comput. Electron. Agric.* 152, 375–384. doi: 10.1016/j.compag.2018.07.029

Tang, Y., Zhou, H., Wang, H., and Zhang, Y. (2023). Fruit detection and positioning technology for a camellia oleifera c. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Syst. Appl.* 211, 118573. doi: 10.1016/j.eswa.2022.118573

Wang, Y., Cai, J., Liu, Y., Chen, X., and Wang, Y. (2022a). Motion-induced error reduction for phase-shifting profilometry with phase probability equalization. *Opt. Lasers Eng.* 156, 107088. doi: 10.1016/j.optlaseng.2022.107088

Wang, Y., Cai, J., Zhang, D., Chen, X., and Wang, Y. (2022b). Nonlinear correction for fringe projection profilometry with shifted-phase histogram equalization. *IEEE Trans. Instrum. Meas* 71, 5005509. doi: 10.1109/TIM.2022.3145361

Wang, H., Lin, Y., Xu, X., Chen, Z., Wu, Z., and Tang, Y. (2022). A study on long-close distance coordination control strategy for litchi picking. *Agronomy-Basel* 12, 1520. doi: 10.3390/agronomy12071520

Wang, Y., Wen, W., Wu, S., Wang, C., Yu, Z., Guo, X., et al. (2018). Maize plant phenotyping: Comparing 3D laser scanning, multi-view stereo reconstruction, and 3D digitizing estimates. *Remote Sens.* 11, 63. doi: 10.3390/rs11010063

Wu, F., Duan, J., Chen, S., Ye, Y., Ai, P., and Yang, Z. (2021). Multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.705021

Xie, L., Xu, Y., Zhang, X., Bao, W., Tong, C., and Shi, B. (2019). A self-calibrated photo-geometric depth camera. *Visual Comput.* 35, 99–108. doi: 10.1007/s00371-018-1507-9

Xue, H., Sun, M., and Liang, Y. (2022). ECANet: Explicit cyclic attention-based network for video saliency prediction. *Neurocomputing* 468, 233–244. doi: 10.1016/j.neucom.2021.10.024

Xu, X., Zhang, L., Yang, J., Liu, C., Xiong, Y., Luo, M., et al. (2021). LiDAR-camera calibration method based on ranging statistical characteristics and improved RANSAC algorithm. *Robot. Auton. Syst.* 141, 103776. doi: 10.1016/j.robot.2021.103776

Xu, X., Zhuge, S., Guan, B., Lin, B., Gan, S., Yang, X., et al. (2022). On-orbit calibration for spaceborne line array camera and LiDAR. *Remote Sens.* 14, 2949. doi: 10.3390/rs14122949

Yang, T., Bai, Z., Li, Z., Feng, N., and Chen, L. (2021). Intelligent vehicle lateral control method based on feedforward. *Actuators* 10, 228. doi: 10.3390/act10090228

Yang, Y., Li, Y., Wen, X., Zhang, G., Ma, Q., Cheng, S., et al. (2022a). An optimal goal point determination algorithm for automatic navigation of agricultural machinery: Improving the tracking accuracy of the pure pursuit algorithm. *Comput. Electron. Agric.* 194, 106760. doi: 10.1016/j.compag.2022.106760

Yang, Z., Ouyang, L., Zhang, Z., Duan, J., Yu, J., and Wang, H. (2022b). Visual navigation path extraction of orchard hard pavement based on scanning method and neural network. *Comput. Electron. Agric.* 197, 106964. doi: 10.1016/j.compag.2022.106964

Yu, J., Ma, L., Tian, M., and Lu, X. (2021). Registration and fusion of UAV LiDAR system sequence images and laser point clouds. *J. Imaging Sci. Technol.* 65, 10501. doi: 10.2352/J.ImagingSci.Technol.2021.65.1.010501

Zhang, Z., Jia, X., Yang, T., Gu, Y., Wang, W., and Chen, L. (2021). Multi-objective optimization of lubricant volume in an ELSD considering thermal effects. *Int. J. Therm. Sci.* 164, 106884. doi: 10.1016/j.ijthermalsci.2021.106884

Zhang, Z., Li, P., Zhao, S., Lv, Z., and An, Y. (2020). An adaptive vision navigation algorithm in agricultural IoT system for smart agricultural robots. *Computers Mater. Continua* 66, 1043–1056. doi: 10.32604/cmc.2020.012517

Zhou, Z., Song, Z., Fu, L., Gao, F., Li, R., and Cui, Y. (2020). Real-time kiwifruit detection in orchard using deep learning on android (TM) smartphones for yield estimation. *Comput. Electron. Agric.* 179, 105856. doi: 10.1016/j.compag.2020.105856

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 December 2022
DOI 10.3389/fpls.2022.1075856

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Jianjun Du,
Beijing Research Center for Information Technology in Agriculture, China
Zipeng Zhang,
Xinjiang University, China

*CORRESPONDENCE
Bing Zhang
zhangbing@aircas.ac.cn

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 21 October 2022
ACCEPTED 28 November 2022
PUBLISHED 21 December 2022

CITATION
Hu J, Zhang B, Peng D, Yu R, Liu Y, Xiao C, Li C, Dong T, Fang M, Ye H, Huang W, Lin B, Wang M, Cheng E and Yang S (2022) Estimation of wheat tiller density using remote sensing data and machine learning methods. *Front. Plant Sci.* 13:1075856. doi: 10.3389/fpls.2022.1075856

COPYRIGHT
© 2022 Hu, Zhang, Peng, Yu, Liu, Xiao, Li, Dong, Fang, Ye, Huang, Lin, Wang, Cheng and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Estimation of wheat tiller density using remote sensing data and machine learning methods

Jinkang Hu[1,2], Bing Zhang[1,2]*, Dailiang Peng[1,3], Ruyi Yu[1], Yao Liu[4], Chenchao Xiao[4], Cunjun Li[5], Tao Dong[6], Moren Fang[7], Huichun Ye[1,3], Wenjiang Huang[1,3], Binbin Lin[8], Mengmeng Wang[9], Enhui Cheng[1,2] and Songlin Yang[1,2]

[1]Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, [2]College of Resource and Environment, University of Chinese Academy of Sciences, Beijing, China, [3]International Research Center of Big Data for Sustainable Development Goals, Beijing, China, [4]Land Satellite Remote Sensing Application Center, Ministry of Natural Resources of China, Beijing, China, [5]Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, [6]Aerospace ShuWei High Tech. Co., Ltd., Beijing, China, [7]Beijing Azup Scientific Co., Ltd., Beijing, China, [8]Department of Geography, Texas A&M University, TX, United States, [9]School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan, China

The tiller density is a key agronomic trait of winter wheat that is essential to field management and yield estimation. The traditional method of obtaining the wheat tiller density is based on manual counting, which is inefficient and error prone. In this study, we established machine learning models to estimate the wheat tiller density in the field using hyperspectral and multispectral remote sensing data. The results showed that the vegetation indices related to vegetation cover and leaf area index are more suitable for tiller density estimation. The optimal mean relative error for hyperspectral data was 5.46%, indicating that the results were more accurate than those for multispectral data, which had a mean relative error of 7.71%. The gradient boosted regression tree (GBRT) and random forest (RF) methods gave the best estimation accuracy when the number of samples was less than around 140 and greater than around 140, respectively. The results of this study support the extension of the tested methods to the large-scale monitoring of tiller density based on remote sensing data.

KEYWORDS

winter wheat, tiller density, UAV hyperspectral, vegetation index, random forest, gradient boosted regression trees

1 Introduction

Wheat is one of the world's most important food crops and provides food for more than half of the world's population (Grassini et al., 2013; Blackie, 2016). With the world population expected to reach 9 billion by 2050, demand for wheat is expected to increase by 60%–110% (Godfray et al., 2010; Tilman et al., 2011; Ray et al., 2013). To meet this demand, annual wheat yield increases must rise from the current value of less than 1% to at least 1.6% (Tilman et al., 2011; Ray et al., 2013). Wheat's yield potential depends on the tiller density at the tillering stage (Elsayed et al., 2018) and, under normal or high-density sowing scenarios, tillers produced in winter wheat from fall until the beginning of January of the following year constitute more than 87% of the final yield (Tilley et al., 2019). The tiller density is also closely related to the nitrogen status of winter wheat (Elsayed et al., 2018). Therefore, accurate, efficient, and real-time knowledge of the tiller density during the tillering stage of winter wheat is important for improving nitrogen fertilization management, obtaining an optimal seed yield, and implementing sustainable agricultural practices (Cheng, 2020).

The tiller density refers to the number of tillers of winter wheat contained in a unit area (e.g., 1 m^2). Currently, the most common method for measuring the tiller density is manual counting, which is extremely time-consuming and inefficient, limited by human error, and lacking in timeliness and accuracy (Scotford and Miller, 2004). Remote sensing provides an alternative method due to its ability to provide quantitative biophysical parameter data for vegetation in a non-contact and non-destructive manner (Zenkl et al., 2021). Remote sensing estimation methods of tiller density in the literature can be generally classified into two types: (1) image segmentation models and (2) spectral feature models. Both 2D and 3D image segmentation models are available: the 2D approaches are based on 2D RGB images taken by handheld cameras or unmanned aerial vehicles (UAVs) and make use of methods such as manually designed features (Liu et al., 2016; Liu et al., 2017; Liu et al., 2018) or machine learning (Jin et al., 2017) to segment leaf image elements so that the tiller density can be estimated under field conditions in sample plots. These methods require a high image resolution (ground sampling distance< 0.5 mm). In the 3D approaches, point clouds of wheat are obtained with the help of remote sensing techniques such as LIDAR, and the tiller number is estimated by clustering (Roth et al., 2020; Fang et al., 2020). This can be severely affected by wind and shading between wheat leaves and cause the tiller number to be underestimated (Fang et al., 2020). Spectral characterization models, in contrast, establish a regression between the tiller density and vegetation indices (VIs) to estimate the tiller density (Flowers et al., 2001; Flowers et al., 2003; Scotford and Miller, 2004; Phillips et al., 2004; Wu et al., 2011; Wu et al., 2022). Most regression models use linear, a few use non-linear ones such as exponential regression. Results show that VIs are reliable indicators for estimating the wheat tiller density in the field; but the relative error was above 20% and could not meet the 10% accuracy required for the application (Liu et al., 2017).

Most current studies of wheat tiller density or tiller number are based on RGB images acquired on the ground or using UAVs; the tiller density is then estimated using image segmentation, which constitutes a source of point data and cannot be used to estimate the tiller density of the plot as a whole; however, it cannot accurately reflect the spatial variation in the density within and between plots. Details of this spatial variation can only be visualized by using a spatial interpolation algorithm and the values of the wheat tiller density that have been obtained, which are subject to errors caused by spatial heterogeneity. In addition, in the case of larger areas, there are difficulties in obtaining UAV data. Developments in high-resolution satellite remote sensing are helping this situation: in particular, spectral feature models can be used to estimate the wheat tiller density on a pixel-by-pixel basis. Therefore, the actual number of tillers of winter wheat can be estimated by using high-resolution satellite images acquired in late fall and early winter based on a small number of measured tillers (Miller and Adkins, 2021); maps showing the spatial distribution of the tiller density can then be obtained.

Traditional methods of inverting crop physicochemical parameters are mainly based on parametric regression of a single vegetation index (VI) as a variable (Verrelst et al., 2015), which is widely used to estimate crop parameters and monitor crop conditions (Bahrami et al., 2021), is used to establish regression relationships. Such methods tend to be very sensitive to noise (Danner et al., 2021) and are suitable for estimating equations corresponding to different linear or exponential relationships (Liang et al., 2015). However, complex and strongly nonlinear relationships exist between biophysical and biochemical parameters and reflectance spectra that cannot be accurately simulated by these parametric models (Liang et al., 2015); also, these models cannot be transferred to other sites with different vegetation or applied to data acquired using other types of sensors or under different conditions (Lu and He, 2019). However, nonparametric linear and nonlinear regression methods have been developed to overcome these deficiencies. In particular, machine learning (ML) regression algorithms have evolved rapidly in recent decades due to their ability to mine and understand information deep within datasets and have been shown to reliably solve nonlinear problems (Camps-Valls et al., 2018). Because of their ability to obtain crop physical and chemical parameters and satellite reflectances, nonlinear modeling of the relationship between physicochemical parameters and satellite reflectance spectra is increasingly applied in combination with remote sensing techniques for crop growth monitoring (Rehman et al., 2019; Zhang et al., 2019; Zha et al., 2020; Machwitz et al., 2021). It is common practice to extract multiple vegetation indices with different effects from spectral information and filter the most relevant vegetation indices to the target physicochemical parameters by using feature engineering or

feature selection (Danner et al., 2021) as the input to train machine learning regression models (e.g., support vector regression (SVR), Gaussian process regression (GPR), random forest (RF), and gradient boosted regression trees (GBRT)). The model with the highest estimation accuracy is then obtained by optimizing and adjusting the model hyperparameters and the cross-validation results. Generally, the number of filtered features does not exceed 15%–20% of the total number of field measurement samples, which means that the risk of overfitting can largely be avoided (Thenkabail et al., 2000). Machine learning methods have evolved as reliable methods of learning nonlinear relationships because they require less parameterization, are implemented at various spatial and temporal scales, and are more robust and covariant to noisy features, small training sizes, and large numbers of dimensions (Verrelst et al., 2012; Liang et al., 2015; Houborg and McCabe, 2018). These methods have been widely used for estimating various biophysical parameters such as the leaf area index (Duan et al., 2019; Tao et al., 2020), vegetation cover (Niu et al., 2021; Yu et al., 2021), biomass (Yue et al., 2019; Tao et al., 2020), Canopy chlorophyll content (Jiao et al., 2021) and the leaf tilth distribution (Zou et al., 2022). However, few studies have been conducted to estimate the tiller density of winter wheat.

Therefore, in this study, models for estimating the tiller density based on multiple vegetation indices using machine learning methods were established. Results with higher accuracy than those obtained in previous research were achieved. Corresponding spatial distribution maps based on different types of remote sensing data (including hyperspectral and multispectral data) were also obtained. It was verified that a machine learning model for estimating the winter wheat tiller density based only on plot-scale samples can be extended to the county scale. In this paper, the use of digital imagery instead of manual counts to determine tiller density as a way of obtaining ground truth data that is less time-consuming and laborious is considered.

2 Materials and methods

2.1 Field experiments, measurements, and data processing

The ground experiments on which this study was based were conducted at two sites near Beijing, China (Figure 1): the Xiaotangshan National Precision Agriculture Research Center (40.10°N, 116.26°E) and Xiongan (38°43′–39°10′N, 115°38′–116°20′E).

The experiments included making the following observations.

① The tiller density was measured by manually counting the number of tillers in a 0.5 m × 0.5 m area around each sampling point. Vertical digital photographs of the same areas were also taken at a 1:1 scale, corresponding to the ground dimensions of 0.5 m × 0.5 m. These photographs

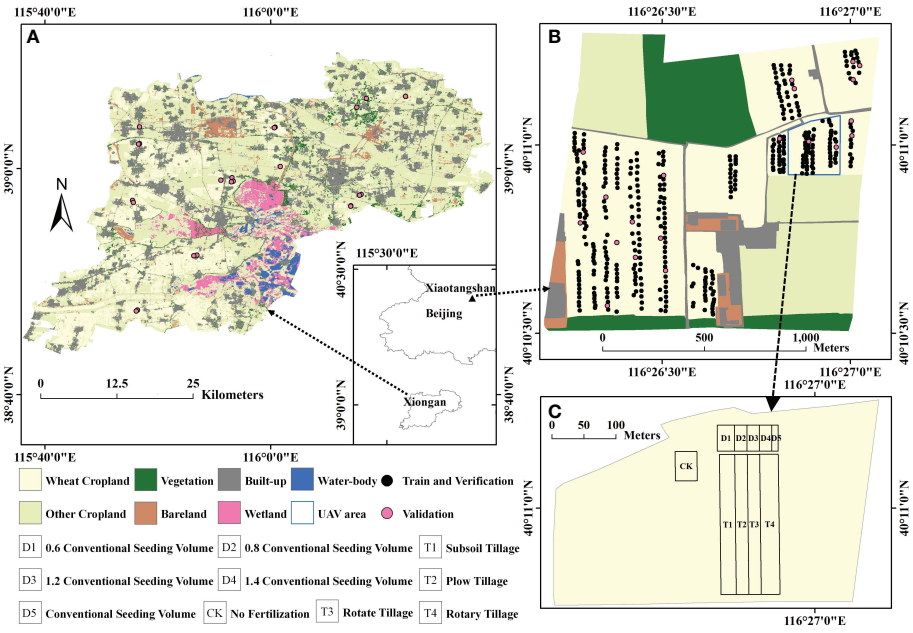


FIGURE 1
Geographical location of the study sites: the experiments were conducted at **(A)** Xiongan in November 2020, **(B)** Xiaotangshan in November and early December 2020, and **(C)** different application scenario settings at Xiaotangshan.

were used for training the deep-learning model that was to be used to extract the tiller density from the photographs.

② Hyperspectral data were acquired using a Cubert S185 image hyperspectral sensor carried by a DJI M300 UAV at an altitude of 40 m on November 23, 2020, at the Xiaotangshan study site. (The area over which these images were acquired is shown as the blue box in Figure 1B). The data were processed to give the ground reflectance in a total of 125 bands within the wavelength range 450–950 nm with a sampling interval of 4 nm. The ground sample distance (GSD) was resampled from 1 cm to 0.5 m to correspond to the imaged area using the nearest neighbor method.

③ The multispectral data used in this study consisted of Sentinel-2 A/B L1C-level satellite data that covered the study areas shown in Figure 1. These data were downloaded from the European Space Agency website (https://scihub.copernicus.eu/dhus/#/home). The Level-2A bottom-of-atmosphere (BOA) reflectance product corresponding to four bands (bands 2, 3, 4, and 8) in the visible and near-infrared range was then obtained by applying the Sen2cor atmospheric correction module provided by ESA to correct for the effects of the atmosphere. For Xiongan, one scene of data from November 2020 was obtained; for Xiaotangshan, four scenes from the period from November to December 2020 were obtained. Further details of the experiments that were carried out at the two study sites are given in Table 1. Mid-to-late November and December were chosen for the experiment because winter wheat fertility had already entered the overwintering season at Xiaotangshan and Xiongan, tillering had ceased, and tiller density was almost unchanged during the growth period. The satellite data were also chosen at this time to correspond with the ground experiment time.

$$\mathbf{BIC} = k\ln(n) - 2\ln(L) \quad (1)$$

where k is the number of model parameters, n is the number of samples, and L is the likelihood function. The BIC criterion is frequently employed as an evaluation criterion for model selection and can effectively circumvent issues that result from models being too complex due to their high accuracy. The BIC criterion also successfully prevents the selection of too many variables when there are too many dimensions and not enough samples. As a result, the minimum BIC value principle—which states that the fewest features carry the greatest information—is applied when choosing variables. In this case, based on the

TABLE 1 Summary of the experiments performed at Xiaotangshan and Xiongan.

Experiment no.	Month (s) and year	Experiment details
Xiaotangshan study site: training, verification, and validation		
1	Nov 2020	Within the area corresponding to the blue box in Figure 1B, observations were made using the ground-based S185 UAV hyperspectral sensor. These observations were made on the same date that the Sentinel-2 satellite data were acquired (November 23). Manually measured tiller density values were collected from 60 points; digital photographs of the same points were taken. (50 of these points were randomly selected for subsequent use as training points; 10 points were used for verifying the accuracy of the estimation model). Ten different application scenarios were set up based on different seeding volumes, different tillage methods, and different fertilization conditions to simulate the tiller density under different scenarios, with specific settings referring to the notes in the lower left corner of Figure 1 and spatial distribution as shown in Figure 1C.
2	Nov and Dec 2020	Within the area of the winter wheat crop marked in Figure 1B, the tiller density value was measured at more than 400 points (including the 60 points from experiment 1). Digital photographs were taken of the same points. These data were used for training and verifying the tiller density estimation model for the entire Xiaotangshan wheat growing area; data from 20 points were used for validating subsequent observations. Four Sentinel-2 scenes were acquired during the period of these field measurements (November 16 to December 10). (Note: The field measurements were carried out four separate times to coincide with the satellite transits, but due to weather conditions, the field experiments could not be performed at the same time as the Sentinel-2 transits; however, the time difference was not more than two days on any occasion.)
Xiongan study site: validation		
3	Nov 2020	Within the area of the winter wheat crop area marked in Figure 1A, manual measurements of the tiller density were made at 23 points (the pink points in Figure 1A) during the period November 7 to November 9. These data were subsequently used for verifying the accuracy of the extended model that was developed in experiment 2. Sentinel-2 satellite data from November 8 were also acquired to coincide with the time of this experiment.

Before constructing the estimation model based on the vegetation indices and tiller density values, 12 vegetation indices (Table 2) were first considered. These indices were based on the vegetation structure (e.g., the leaf area index, canopy depression, green biomass, or species) or on biochemical parameters (e.g., chlorophyll or other pigments and nitrogen), which were calculated based on the remote sensing data obtained in the previous treatments. The 12 vegetation indices chosen were all broad-band vegetation indices with no hyperspectral vegetation indices to compare the accuracy of hyperspectral and multispectral data results. The tiller density estimation model was constructed using the red, green, blue, and near-infrared bands of the hyperspectral data that had been shown to have the highest correlation with the tiller density. Calculations for Table 2's vegetation indices utilized hyperspectral data at wavelengths of 458 nm, 492 nm, 750 nm, and 740 nm, respectively.

TABLE 2 Summary of vegetation indexes selected for use in this study.

Vegetation Index	Definition	Features	Reference
Vegetation structure			
DVI (Difference Vegetation Index)	$\rho_{nir}-\rho_{red}$	Estimated vegetation leaf area index	Clevers, 1986
EVI (Enhanced Vegetation Index)	$2.5\times(\rho_{nir}-\rho_{red})/(\rho_{nir}+6\times\rho_{red}-7.5\times\rho_{blue}+1)$	Capable of improving the sensitivity of estimates of the vegetation leaf area index, biomass, and water content in areas with high biomass	Huete et al., 2002
MSR (Modified Red-Edge Simple Ratio Vegetation Index)	$\frac{\frac{\rho_{nir}}{\rho_{red}}-1}{\sqrt{(\frac{\rho_{nir}}{\rho_{red}}+1)}}$	Linear correlation with vegetation parameters is higher than for the RDVI	Chen, 1996; Haboudane et al., 2004
MTVI (Modified Triangular Vegetation Index)	$1.2\times(1.2\times(\rho_{nir}-\rho_{green})-2.5\times(\rho_{red}-\rho_{green})$	Sensitive to the leaf area index; suitable for leaf area index estimation	Haboudane et al., 2004
NDVI (Normalized Difference Vegetation Index)	$\frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+\rho_{red}}$	Better response to changes in green biomass; gives improved results for medium- and low-density vegetation	Rouse et al., 1974
NGRDI (Normalized Green Red Difference Vegetation Index)	$\frac{\rho_{green}-\rho_{red}}{\rho_{green}+\rho_{red}}$	Minimizes atmospheric effects on estimates of green vegetation	Gitelson et al., 2002
SR (Simple Ratio Vegetation Index)	$\frac{\rho_{nir}}{\rho_{red}}$	Similar to the NDVI	Jordan, 1969
OSAVI (Optimized Soil-Adjusted Vegetation Index)	$\frac{(1+0.16)(\rho_{nir}-\rho_{red})}{\rho_{nir}+\rho_{red}+0.16}$	More sensitive than the SAVI at the canopy scale and more suitable for agricultural applications	Rondeaux et al., 1996
RGD (Red Green Difference Vegetation Index)	$\rho_{red}-\rho_{green}$	Can be used for estimating vegetation cover	Sanjerehei, 2014
WDRVI (Wide Dynamic Range Vegetation Index)	$\frac{0.1\times\rho_{nir}-\rho_{red}}{0.1\times\rho_{nir}+\rho_{red}}$	Capable of providing better estimation results for the leaf area index, biomass, and vegetation cover than the NDVI	Gitelson, 2004
Biochemical parameters			
MCARI (Modified Chlorophyll Absorption in Reflectance Vegetation Index)	$1.2\times(2.5\times(\rho_{nir}-\rho_{red})-1.3\times(\rho_{nir}-\rho_{green})$	Capable of responding to changes in chlorophyll and estimating the chlorophyll uptake	Daughtry et al., 2000
RDVI (Renormalized Difference Vegetation Index)	$\frac{\rho_{nir}-\rho_{red}}{\sqrt{(\rho_{nir}+\rho_{red})}}$	Capable of quantifying changes in a wide range of chemicals in vegetation; can be applied to a wide range of values of the leaf area index	Roujean and Breon, 1995

These 12 vegetation indices data also needed to be filtered using the Backward Feature Selection (BFS) method with the Bayesian Information Criterion (BIC) (Burnham and Anderson, 2002) as the criterion for removing redundant features before they could be used as inputs to the tiller density estimation model. The BIC was calculated as

criterion function, the feature selection process determined the amount of tiller density information contained from the complete set of vegetation index samples; the redundant vegetation indices were then eliminated one at a time until the final subset of vegetation indices containing the necessary number of features was obtained. The selected vegetation indices were then used as inputs for training the tiller density estimation model.

The tiller density was extracted from the digital images of winter wheat gathered at Xiaotangshan that were described in Table 1. Each image consisted of measurements of the tiller density together with coordinate data. The images were first filtered to remove any blurred images; a total of 2600 JPG images were saved in a 1024 × 1024 × 3 RGB format. The remaining 2400 images were cutted and cropped to an 8:1:1 ratio for later use in training and validation. For transfer learning, PyTorch Hub's DenseNet pre-training model (https://pytorch.org/hub/) was used. For this, the batch size was set to 8 and the learning rate was initially set to 0.01; Adam was chosen as the optimizer, the L2 regularization coefficient was set to 0.00005, the exponential decay rate of first-order moment estimation was set to float between 0.9 and 0.99, and the exponential decay rate of second-order moment estimation is set to 0.999. The tiller density extraction model of digital photos is finally obtained by monitoring MRE for 5 consecutive training rounds without further decline to set Early Stopping to prevent model overfitting. Then using the model to extract the tiller density for subsequent labeling of the relationship between vegetation index and tiller density.

2.2 Method for estimating the tiller density of winter wheat

The filtered vegetation index features were used as the input of the machine learning model. Five classical models were chosen for the machine learning method: Ordinary Least Squares (OLS), Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Regression Tree (GBRT), and Extreme Gradient Boosting (XGBoost). The samples of observed tiller density (the black points shown in Figure 1B) were randomly separated into training and verification sets in the ratio 8:2 for five-fold cross-validation; the labels consisted either of manually measured values of the tiller density or values that

had been extracted from the digital photographs. The hyperparameters of the five models, including n estimators, max depth, min samples split, min samples leaf, and max features, were inputted in dictionary form. And the GridSearchCV method was used to adjust the hyperparameters before the optimal hyperparameter values were output. The best model was selected that gave the highest accuracy when applied to the verification set. The correlation coefficient, r, and mean relative error (MRE) were used to determine an evaluation index that described the accuracy of the tiller density estimation model. The p-value was also used as a measure of the accuracy, and only models with $p< 0.05$ were selected. This helped to guarantee that the results were statistically significant and could minimize overfitting caused by the small sample numbers. The correlation coefficient and MRE were calculated as follows:

$$r = \sqrt{1 - \frac{\sum (y_i - \overline{y_i})^2}{\sum (y_i - \widetilde{y_i})^2}} \quad (2)$$

$$MRE = \frac{1}{m}\sum_{i}^{m}\left|\frac{\overline{y_i} - y_i}{y_i} \times 100\,\%\right| \quad (3)$$

$$t = \frac{r\sqrt{m-2}}{\sqrt{1-r^2}} \quad (4)$$

Here y_i is the predicted value, $\overline{y_i}$ is the true value, and m is the number of samples, t is the t-distribution; the p-value was obtained from the t-distribution corresponding to the correlation coefficient. The correlation coefficient was used to determine the model fitting regression effect: the closer the value of this was to 1, the better the regression effect. The MRE is defined as the average ratio of the absolute error of the measurement to the actual measurement. the smaller the value of this, the better the model. The p-value is a measure of the probability and gives the likelihood of an event occurring: generally $p< 0.05$ means a statistical difference, $p< 0.01$ is a statistically significant difference, and $p< 0.001$ is an extremely significant difference.

3 Results and discussion

3.1 Results of tiller density estimation under different experimental conditions

3.1.1 Tiller density values obtained by different machine learning methods

Based on the UAV hyperspectral data and the Sentinel multispectral data from the same site in Xiaotangshan (marked as the blue box in Figure 1B), the vegetation indices listed in Table 2 were calculated and then filtered. After filtering, the indices MCARI, RDVI, and WDRVI were obtained from the hyperspectral data, and NDVI, DVI, MCARI, MSR, RGD, RVI, and WDRVI were obtained from the multispectral data. Estimation models were then built using different machine learning models based on the manually measured values of the tiller density. Predictions of the tiller density for the same area were then made, and the spatial distribution of these values was obtained, as shown in Figure 2 (The results for the RF and GBRT methods are shown here; the hyperspectral data were resampled to 10 m using the nearest neighbor method to facilitate comparison with the multispectral data.). The estimation results of the hyperspectral image are more compatible with the actual spatial distribution of tiller density than those of multispectral images for the various types of remote sensing data. The estimation results of GBRT are more compatible with the actual spatial distribution of tiller density for the same type of remote sensing data as those of RF (Figure 2).

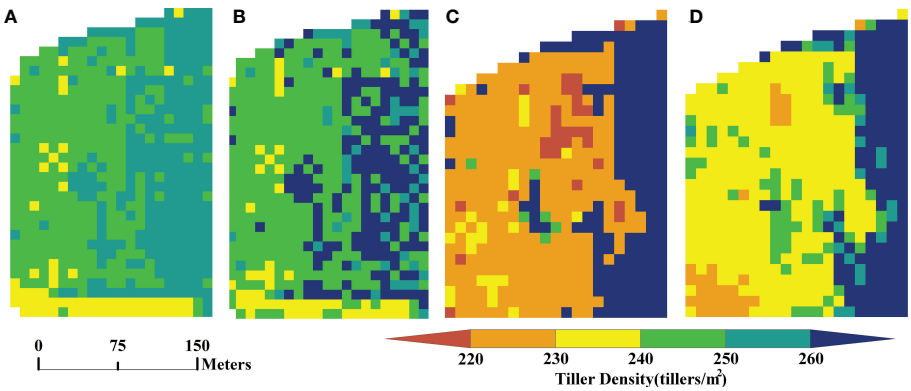


FIGURE 2
Spatial distribution of tiller density estimated using hyperspectral and multispectral data: **(A)** results for RF model applied to hyperspectral data, **(B)** results for GBRT model applied to hyperspectral data, **(C)** results for RF model applied to multispectral data, and **(D)** results for GBRT model applied to multispectral data.

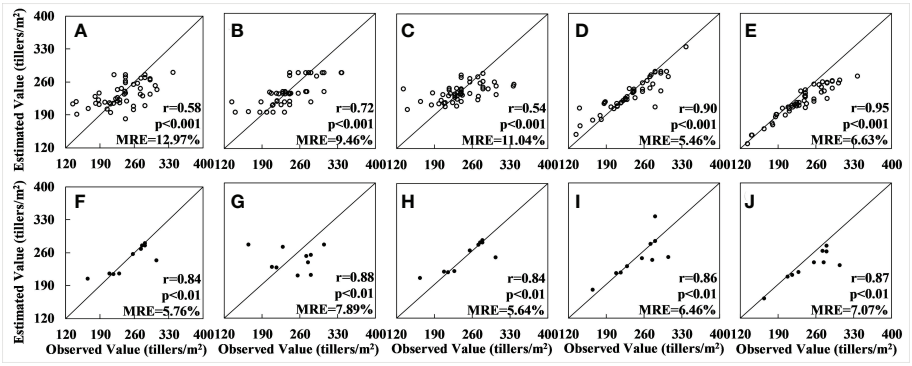


FIGURE 3
Accuracy of tiller density estimates based on the UAV hyperspectral data: **(A–E)** show the regression results for applying, respectively, the OLS, RF, SVM, GBRT, and XGBoost models to the training set; **(F–J)** show the regression results for the applying the same five models to the validation set.

Next, the verification dataset was used to verify the accuracy of the tiller density estimation. The results for the GBRT model were found to have the highest accuracy among the results for the hyperspectral data (r = 0.90 and MRE = 5.46% for the training set (see Figure 3D) and r = 0.86 and MRE = 6.46% for the verification set) (see Figure 3I). The results for the XGBoost model (see Figures 3E, J) had the greatest relative error up to 0.03 compared to those for the GBRT model, and the correlation coefficient for the training set was lower than the GBRT model. A comprehensive analysis also showed that the fitting effect was inferior to that for GBRT. The RF (see Figures 3B, G), SVM (see Figures 3C, H), and OLS (see Figures 3A, F) models performed much worse on the training set than the GBRT. The results for the RF model showed significant overfitting when the sample numbers were minimal because this model uses the average value at the root node as the outcome (see Figures 3B, G).

Among the results for the multispectral data, the results of the GBRT model had the highest accuracy (r = 0.88 and MRE = 7.71% for the training set and r = 0.64 and MRE = 8.95% for the verification set). The XGBoost model results were poorer than those for the GBRT, with a relative error of 0.01–0.015, a lower r-value, and an inferior fitting effect, and the accuracy of the RF, SVM and OLS models was significantly lower than that of the GBRT. In particular, although the OLS method produced results with good accuracy for the training set (r = 0.70, MRE = 9.91%), validation with the verification set produced results that deviated greatly from the observed value. The fitting effect was also very poor, and serious overfitting occurred; the scatter plot for the

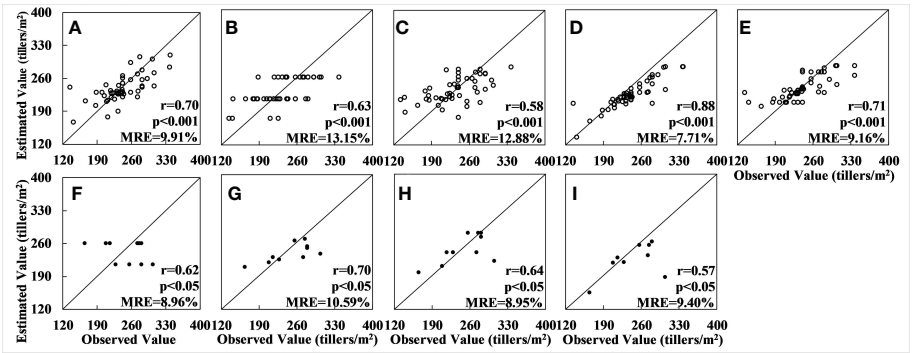


FIGURE 4
Accuracy of tiller density estimates based on Sentinel-2 data: **(A–E)** are the regression results obtained by applying, respectively, the OLS, RF, SVM, GBRT, and XGBoost models to the training set; **(F–I)** are the results obtained by applying, respectively, the RF, SVM, GBRT, and XGBoost models to the validation set. (The results for the OLS model could not be fitted and no results for the accuracy were obtained.).

verification set is therefore not shown in Figure 4. The results for the RF method also showed serious overfitting (see Figures 4B, F).

3.1.2 Tiller density estimates based on different sample numbers

The experimental area was then expanded to include the whole of the wheat crop area at Xiaotangshan base shown in Figure 1B. Based on the Sentinel multispectral data, values of the vegetation indices were again calculated and filtered. The selected vegetation indices were the NDVI, DVI, MCARI, MSR, RGD, RVI, and WDRVI. Tiller density estimation models based on different machine learning models were then built. The spatial distribution of the tiller density was again obtained using these models. Figure 5 shows the results obtained using the RF and GBRT models.

In terms of the spatial distribution, the results obtained using the RF model are more uniform than the other four models and correspond better to the actual situation. The GBRT model results are more random; there are also large differences between neighboring tiller density values in the same region.

The results for the model accuracy obtained using the verification dataset are shown in Figure 6. It can be seen that, in this case, the model with the best accuracy is the RF model: for the training set the results are r = 0.85 and MRE = 10.25%, and for the verification set they are r = 0.66 and MRE = 14.13%. Among the other four models, the GBRT model performed slightly worse than the RF model on both the training and verification sets. The relative errors for the other three models – OLS, SVM, and XGBoost – have increased as a result of the larger number of samples, and the values of the accuracy are significantly lower than for the RF model. (The total number of samples, in this case, was 400; these were divided into training and verification sets using a ratio of 8:2). The experimental area was also larger.

The results show that, for both the hyperspectral and multispectral datasets, when the number of data is less than 140 or so, models which are based on the boosting concept, such as GBRT and XGBoost, work best. Models that are based on the bagging concept, such as RF, perform less well due to the influence of outliers, as this leads to a concentration of values in the results. The SVM model, which maps the data from linear to nonlinear using kernel functions, is also affected by this problem to some extent. The OLS model is completely unsuitable for nonlinear fitting with a large number of features. If the sample number is greater than 140 or so, the RF model outperforms the GBRT model in terms of estimation accuracy because the RF model is sensitive to excessively unstable conditions when the sample numbers are small and cannot effectively reject outliers, resulting in overfitting. In contrast, the serial structure of the GBRT model avoids this situation when the sample numbers are small. When the number of samples increases to more than 140 or so, the RF model performs better due to good noise immunity (see Figure 7).

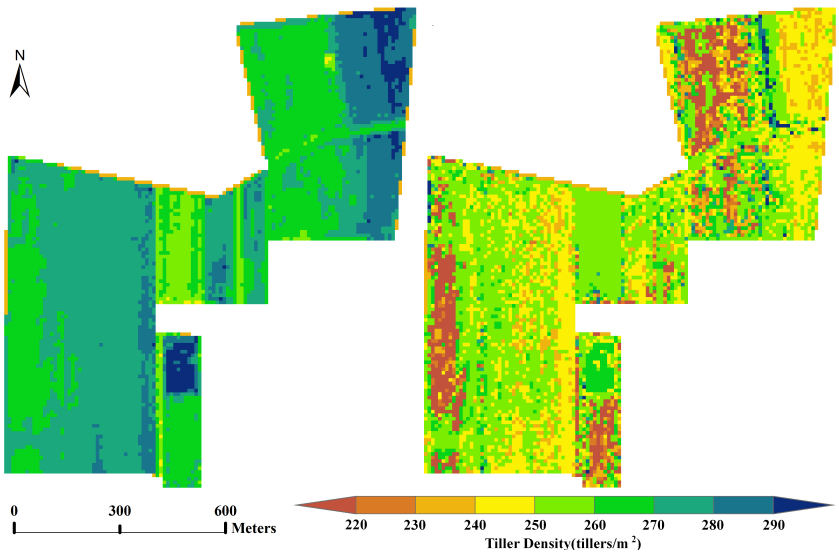


FIGURE 5
Images showing the spatial distribution of tiller density estimates obtained using Sentinel multispectral data of the whole Xiaotangshan wheat growing area: **(A)** RF model results, and **(B)** GBRT model results.

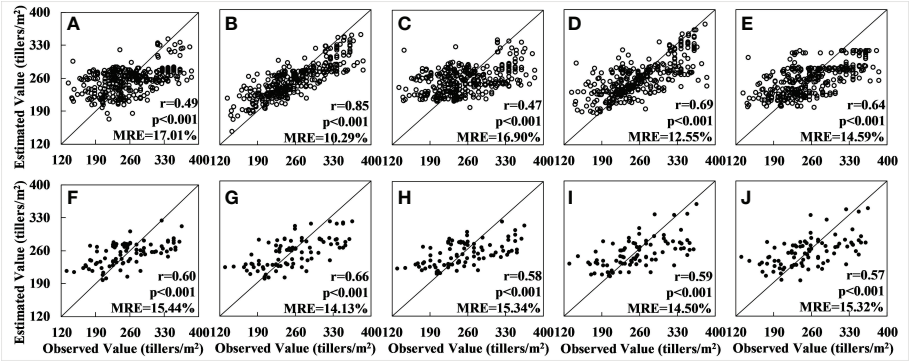


FIGURE 6
Accuracy of tiller density estimates based on Sentinel-2 data for the whole Xiaotangshan study area: **(A–E)** are the regression results for, respectively, the models OLS, RF, SVM, GBRT, and XGBoost for the training set; **(F–I)** are the regression results for the same five models for the validation set.

3.1.3 Tiller density estimates based on different types of samples

In the next experiment, the whole wheat planting area at the Xiaotangshan base was again used. This time, however, the tiller density values were extracted from the digital photos and manually counted data. For the manually counted samples, the vegetation indices MCARI, EVI, RDVI, OSAVI, and DVI were used; the MTVI, RGD, EVI, RDVI, and OSAVI were used for the digital photographs. In Section 3.1.2, it was shown that the RF model is the most accurate when the sample number was greater than 140 or so. Therefore, a tiller density estimation model based on the RF model was built and then validated using the independent validation set (the pink points in Figure 1B). Based on the manually counted values, a value of r = 0.80 was obtained with a relative error of 8.66%; for the values extracted from the digital photographs, the value of r was 0.85 and the relative error was 8.98%. In both cases, p< 0.001, meaning that the results were statistically significant.

As described in Section 3.1.2, if more than around 140 samples were used, the accuracy of the RF model increased. Therefore, the independent validation set was used to validate the tiller density estimation model based on the RF model. The results for both the manually counted values and the digital photograph values were statistically significant (p< 0.001); the relative errors were 8.66% and 8.98%, respectively.

It can be concluded that tiller density extracted from digital photographs can be used in place of manually counted values as the accuracy of the estimates based on the two sets of data was similar. This would increase the effectiveness of sample collection and reduce errors due to subjective human judgment.

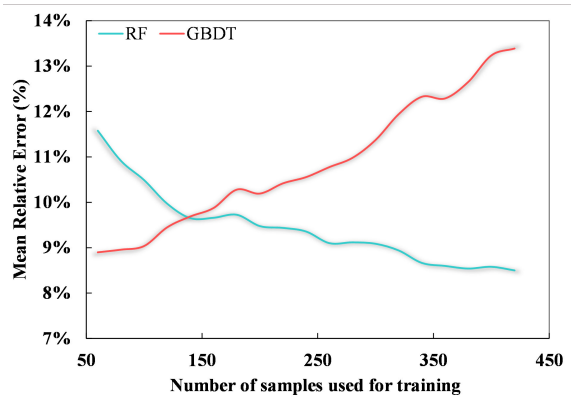


FIGURE 7
The variation in the mean relative error of RF (cyan) and GBRT (red) estimation results with increasing sample size on the validation set.

3.2 Analysis of the results obtained by applying the model to a larger area

In this section, the random forest tiller density estimation model established in Section 3.1.2 utilizing plot-scale sample data from the Xiaotangshan study site was extended to the Xiongan winter wheat crop area for use, and the same seven vegetation indices NDVI, DVI, MCARI, MSR, RGD, RVI, and WDRVI were used to predict the winter wheat tiller density in Xiongan.

The estimated tiller densities obtained in this way are shown in Figure 8. These results distinguish better between different densities than the other four models, and the corresponding tiller densities within the same plot of land are more uniform. Even the boundaries been plots can be approximately identified, which may be because uniform sowing is used for planting in large fields. Most of the estimated values are in the range of 235–275 tillers/m2; values of 240–270 tillers/m2 correspond to about 160,000–

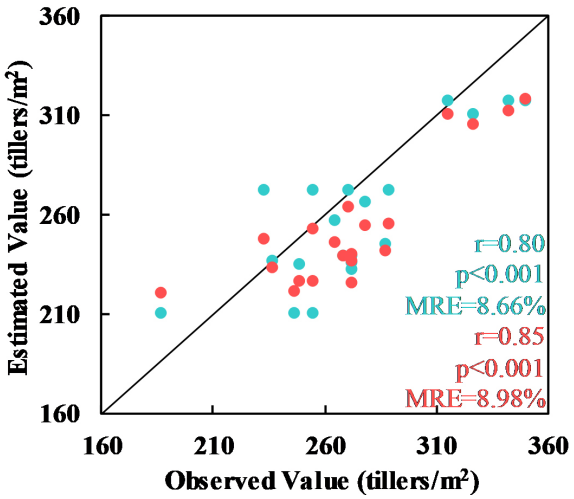


FIGURE 8
Estimated values of the tiller density based on values extracted from manually counted data (cyan) and digital photographs (red).

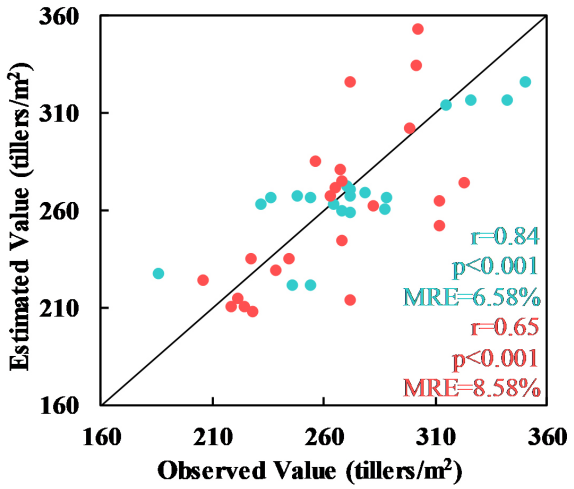


FIGURE 10
Comparison of the tiller density estimates at Xiaotangshan (cyan) and in the Xiongan study area (red) with another validation dataset.

180,000 tillers per acre, which is in agreement with the 120,000–180,000 tillers per acre used when sowing (see Figure 9).

Validation of the Xiongan results using the ground validation points (the pink points in Figure 1A) gave a statistically significant value of r of 0.65 and a relative error of 8.58% with p< 0.001. The same model was also validated as statistically significant using other validation points (the pink points shown in Figure 1B) at the Xiaotangshan research site (r = 0.84, MRE = 6.58%, p< 0.001) (see Figure 10).

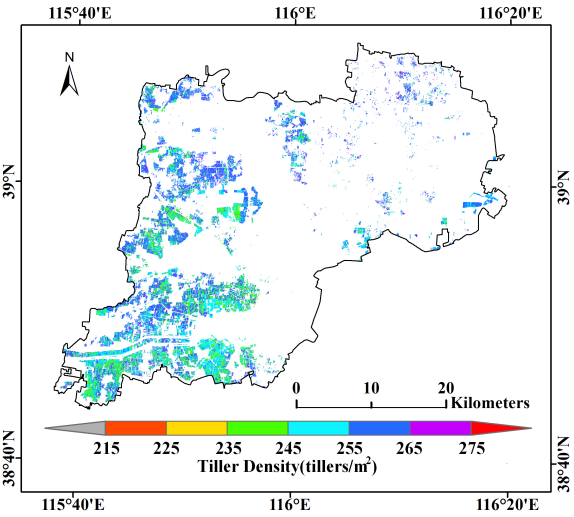


FIGURE 9
Estimated values of the tiller density obtained by extending the RF model to the Xiongan winter wheat crop area.

In conclusion, the validation results of both spatial distribution and ground observed values two ways, demonstrate that the plot-scale tiller density estimation model obtained using data from the Xiaotangshan study site can be extended to the county scale. Although the validation of the results showed that the accuracy of the Xiongan results was lower than that of the Xiaotangshan results, the relative error was still less than 10%, which is sufficient for making estimates of the tiller density of winter what over large areas based on satellite remote sensing data.

3.3 Comparison and analysis with other tiller density estimation methods

The tiller density estimation method developed in this study was primarily based on the spectral features of remote sensing data, and quantitative estimates that benefitted from the advantages of machine learning methods as applied to nonlinear regression were obtained using multiple vegetation indices after feature selection. In contrast, the traditional approach to tiller density estimation is generally based on the high degree of correlation between NIR bands and the tiller density and uses regression based on a single vegetation index (Flowers et al., 2001; Flowers et al., 2003). The results obtained in this way are often biased (relative error > 20%) (Scotford and Miller, 2004), and the generalizability of the method is poor due to the limited applicability of the chosen vegetation index. For example, the NDVI does not fully reflect the wheat tiller density in situations where the leaf area index is particularly high or low or where the amount of cover is high; the sensitivity of the RVI

decreases significantly when the vegetation cover is below 50% and thus does not fully reflect the number of wheat tillers (Wu et al., 2022). To address these problems, in this study, multiple vegetation indices were used to complement each other to meet the accuracy requirements of precision agriculture.

Whether based on 2D RGB images or 3D point cloud data, the results of tiller density estimation based on an image segmentation model are susceptible to the influence of the wind as well as the lighting conditions (Roth et al., 2020). The resolution of the 2D RGB images also needs to be high (Jin et al., 2017). Both types of data are mainly captured by UAVs or handheld cameras. This can lead to errors associated with the chosen sampling location selection as a result of spatial heterogeneity (Liu et al., 2017), meaning that the acquired data can only be applied at the scale of an individual plot of land.

It has been shown that our method meets the accuracy requirements to estimate tiller density using high-resolution remote sensing data and can be used to obtain complete maps of the spatial distribution of the tiller density within an individual plot, which is something that methods based on image segmentation cannot do. It has also been shown that the proposed method can be extended to larger-scale tiller density estimation and monitoring, thus taking full advantage of the ability of remote sensing to quickly provide data covering large areas and potentially aiding more accurate fertilizer application and yield estimation.

In addition, this study has also provided a preliminary demonstration of the feasibility of using digital photographs instead of manual counting. In the future, the use of accurate values of the tiller density extracted from digital photographs could be extended to larger samples to achieve low-cost estimates of the spatial distribution of the tiller density at large spatial scales, something which has not been considered in previous studies.

3.4 Analysis of factors affecting the accuracy of the tiller density estimation

3.4.1 Relationship between the tiller density and the type of remote sensing data

According to the results presented in Section 3.1.1, for all five machine learning methods that were tested, the tiller density estimates based on the hyperspectral data were more accurate than those based on the multispectral data. This was primarily due to the hyperspectral data's high spectral resolution and the large number of bands, which allowed vegetation indices to be constructed using the bands that were correlated most strongly with the tiller density. Other vegetation indices could be used as well as the narrow-band indices that were used in this study (Borengasser et al., 2007). However, the central wavelengths of the bands of the Sentinel data were marginally less well correlated with the tiller density than the hyperspectral data were, which led to the lower accuracy of the results for the multispectral data.

3.4.2 Relationship between the tiller density and vegetation indexes

As described in Section 3.1, the features selected from hyperspectral data were the MCARI, RDVI, and WDRVI, the features selected from the Sentinel data were the NDVI, DVI, MCARI, MSR, RGD, SR, and WDRVI, the features selected from the manual measurements were the MCARI, EVI, RDVI, OSAVI, and DVI, and the features selected from the values extracted from the digital photographs were the MTVI, RGD, EVI, RDVI, and OSAVI. These vegetation indices are mainly related to vegetation parameters such as chlorophyll content, leaf area index, vegetation cover, and aboveground biomass. The vegetation indices related to the leaf area index and vegetation cover are the most frequent, which is a crucial measure of crop growth (Xing et al., 2021). The strong correlation between the vegetation indices that reflect the chlorophyll content of the wheat canopy surface (such as the NDVI) and the tiller density has been demonstrated in previous studies, (Flowers et al., 2001; Flowers et al., 2003). This is reflected in how the tiller density affects the value of the leaf area index and the canopy density (Bates et al., 2021). Changes in the canopy density also cause changes in the vegetation cover, which means that the tiller density can be estimated from the vegetation cover: this is consistent with the observation of a strong correlation between the vegetation cover and plant density in wheat. (Wang et al., 2020; Wu et al., 2022).

Although these vegetation indices may be linearly correlated with each other, it is still possible to estimate the tiller density from them thanks to the ability of machine learning to handle covariance problems (Liang et al., 2015). Therefore, consideration should be given to the use of vegetation indices related to vegetation cover when selecting which indices to use for estimating tiller density.

3.4.3 Relationship between tiller density and texture

Texture, another important class of features that can be used for the inversion of vegetation parameters, is widely used with machine learning inversion methods such as AGB (Yue et al., 2019). In this study, based on the grayscale coevolution matrix (GLCM), we also attempted to calculate eight image texture features for four different bands: mean, variance, homogeneity, contrast, dissimilarity, entropy, second-order moments, and correlation (Yue et al., 2019). Together with the vegetation index, these features were filtered based on the BIC criterion using the BFS method; it was found that the vegetation index

contained the most information about the tiller density while the image texture features contained little information. For both the UAV and Sentinel-2 data, the EVI and the other vegetation indices that made use of the NIR bands ranked highly in terms of tiller density information content, which is consistent with the findings of Flowers et al. (Flowers et al., 2001; Flowers et al., 2003; Scotford and Miller, 2004). The reason for this may be that the tiller density at tillering stage is a relatively microscopic feature: the individual tillers overlap each other, which makes them difficult to distinguish with the naked eye, and an extremely high spatial resolution (e.g., 0.02 cm) is required to extract information using machine vision methods (Liu et al., 2018). The resolution of the data used in this study did not meet this requirement. However, as the tiller density increases, the canopy density and the amount of cover change, which also affects the spectral features (the reflectance in the near-infrared band increases). The vegetation indices can amplify this effect, thus making more tiller density information available and better estimates possible.

4 Conclusion

In this study, we attempted to estimate the tiller density of winter wheat at the tillering stage based on a combination of multiple remotely sensed vegetation indices and using machine learning models.

(1) Under all experimental conditions, the relative error in the estimates of the tiller density was in the range of 5.46%–12.97% for the hyperspectral data and 7.71%–13.15% for the multispectral data. The estimates based on the hyperspectral data were thus more accurate, and in both cases, the relative error was less than 10%, which is the usual level of accuracy required.

Based on the results of this study, tiller density can be extracted from digital images instead of by manual counting during ground sampling as the results for the tiller density obtained in this way were just as accurate as those based on the manual method.

(2) The application of this machine learning model for estimating the tiller density of winter wheat based on plot-scale samples could be extended to the county scale and still meet the requirement of having a relative error of less than 10% although the results may be affected by the spatial heterogeneity of the wheat.

Among the different methods that were tested, the random forest and gradient boosting tree methods gave the most accurate results. The gradient boosting tree is most suitable for sample numbers less than around 140; the random forest is suitable for sample numbers greater than around 140 or with outliers.

(3) Vegetation indices associated with the vegetation cover and leaf area index are suitable for use as features for estimating the winter wheat tiller density. The texture features in remote sensing imagery contain almost no information on the winter wheat tiller density and are hence not a suitable basis for making estimates of the tiller density.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

The experiment was mainly conceived and designed by BZ. JH, RY, YL, CX, CL, TD, MF, HY, WH, BL, EC, and SY performed the experiments. BZ, JH, and DP analyzed the data. The algorithm development was mainly accomplished by BZ, JH, and DP. JH wrote the manuscript and BZ made very significant revisions. DP and MW also read and improved the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China under Grant 42030111 and the National Key R&D Program of China under Grant 2021YFB3900503.

Conflict of interest

Author TD was employed by Aerospace ShuWei High Tech. Co., Ltd. and author MF was employed by Beijing Azup Scientific Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bahrami, H., Homayouni, S., Safari, A., Mirzaei, S., Mahdianpari, M., and Reisi-Gahrouei, O. (2021). Deep learning-based estimation of crop biophysical parameters using multi-source and multi-temporal remote sensing observations. *Agronomy* 11 (7), 1363.

Bates, J. S., Montzka, C., Schmidt, M., and Jonard, F. (2021). Estimating canopy density parameters time-series for winter wheat using UAS mounted lidar. *Remote Sens.* 13 (4), 710.

Blackie, M. (2016). *Save and grow in practice: a guide to sustainable cereal production. maize, rice, wheat; a guide to sustainable production FAO*. Eds. T. Reeves, G. Thomas and G. Ramsay (Rome: Springer).

Borengasser, M., Hungate, W. S., and Watkins, R. (2007). *Hyperspectral remote sensing: principles and applications* (CRC press).

Burnham, K.P., and Anderson, D.R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. (New York: Springer)

Camps-Valls, G., Martino, L., Svendsen, D. H., Campos-Taberner, M., Muñoz-Marí, J., Laparra, V., et al. (2018). Physics-aware Gaussian processes in remote sensing. *Appl. Soft Computing* 68, 69–82.

Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* 22 (3), 229–242.

Cheng, H. (2020). Future earth and sustainable developments. *Innovation* 1 (3).

Clevers, J. W. (1986). The application of a vegetation index in correcting the infrared reflectance for soil background. In Remote sensing for ressources development and environmental management. *International symposium*. 7, pp. 221–226

Danner, M., Berger, K., Wocher, M., Mauser, W., and Hank, T. (2021). Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. *ISPRS J. Photogrammetry Remote Sens.* 173, 278–296.

Daughtry, C. S., Walthall, C. L., Kim, M. S., De Colstoun, E. B., McMurtrey, J. E. III, et al (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote sensing of Environment*, 74 (2), 229–239

Duan, B., Liu, Y., Gong, Y., Peng, Y., Wu, X., Zhu, R., et al. (2019). Remote estimation of rice LAI based on Fourier spectrum texture from UAV image. *Plant Methods* 15 (1), 1–12.

Elsayed, S., Barmeier, G., and Schmidhalter, U. (2018). Passive reflectance sensing and digital image analysis allows for assessing the biomass and nitrogen status of wheat in early and late tillering stages. *Front. Plant Sci.* 9, 1478.

Fang, Y., Qiu, X., Guo, T., Wang, Y., Cheng, T., Zhu, Y., et al. (2020). An automatic method for counting wheat tiller number in the field with terrestrial LiDAR. *Plant Methods* 16 (1), 1–14.

Flowers, M., Weisz, R., and Heiniger, R. (2001). Remote sensing of winter wheat tiller density for early nitrogen application decisions. *Agron. J.* 93 (4), 783–789.

Flowers, M., Weisz, R., Heiniger, R., Tarleton, B., and Meijer, A. (2003). Field validation of a remote sensing technique for early nitrogen application decisions in wheat. *Agron. J.* 95 (1), 167–176.

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.* 161 (2), 165–173.

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80 (1), 76–87.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *science* 327 (5967), 812–818. doi: 10.1126/science.1185383

Grassini, P., Eskridge, K. M., and Cassman, K. G. (2013). Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* 4 (1), 1–11.

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 90 (3), 337–352.

Houborg, R., and McCabe, M. F. (2018). A hybrid training approach for leaf area index estimation *via* cubist and random forests machine-learning. *ISPRS J. Photogrammetry Remote Sens.* 135, 173–188.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83 (1-2), 195–213.

Jiao, Q., Sun, Q., Zhang, B., Huang, W., Ye, H., Zhang, Z., et al. (2021). A random forest algorithm for retrieving canopy chlorophyll content of wheat and soybean trained with PROSAIL simulations using adjusted average leaf angle. *Remote Sens.* 14 (1), 98.

Jin, X., Liu, S., Baret, F., Hemerlé, M., and Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sens. Environ.* 198, 105–114.

Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 50 (4), 663–666.

Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S., et al. (2015). Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sens. Environ.* 165, 123–134.

Liu, S., Baret, F., Andrieu, B., Burger, P., and Hemmerlé, M. (2017). Estimation of wheat plant density at early stages using high resolution imagery. *Front. Plant Sci.* 8, 739.

Liu, T., Wu, W., Chen, W., Sun, C., Zhu, X., and Guo, W. (2016). Automated image-processing for counting seedlings in a wheat field. *Precis. Agric.* 17 (4), 392–406.

Liu, T., Yang, T., Li, C., Li, R., Wu, W., Zhong, X., et al. (2018). A method to calculate the number of wheat seedlings in the 1st to the 3rd leaf growth stages. *Plant Methods* 14 (1), 1–14.

Lu, B., and He, Y. (2019). Evaluating empirical regression, machine learning, and radiative transfer modelling for estimating vegetation chlorophyll content using bi-seasonal hyperspectral images. *Remote Sens.* 11 (17), 1979.

Machwitz, M., Pieruschka, R., Berger, K., Schlerf, M., Aasen, H., Fahrner, S., et al. (2021). Bridging the gap between remote sensing and plant phenotyping–challenges and opportunities for the next generation of sustainable agriculture. *Front. Plant Sci.* 12, 2334.

Miller, J. O., and Adkins, J. (2021). Monitoring winter wheat growth at different heights using aerial imagery. *Agron. J.* 113 (2), 1586–1595.

Niu, Y., Han, W., Zhang, H., Zhang, L., and Chen, H. (2021). Estimating fractional vegetation cover of maize under water stress from UAV multispectral imagery using machine learning algorithms. *Comput. Electron. Agric.* 189, 106414.

Phillips, S., Keahey, D., Warren, J., and Mullins, G. (2004). Estimating winter wheat tiller density using spectral reflectance sensors for early-spring, variable-rate nitrogen applications. *Agron. J.* 96 (3), 591–600.

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS One* 8 (6), e66428.

Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., and Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Comput. Electron. Agric.* 156, 585–605.

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55 (2), 95–107.

Roth, L., Camenzind, M., Aasen, H., Kronenberg, L., Barendregt, C., Camp, K.-H., et al. (2020). Repeated multiview imaging for estimating seedling tiller counts of wheat genotypes using drones. *Plant Phenomics* 2020.

Roujean, J.-L., and Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* 51 (3), 375–384.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the great plains with erts. *NASA Special Publ.* 351.

Sanjerehei, M. M. (2014). Assessment of spectral vegetation indices for estimating vegetation cover in arid and semiarid shrublands. *Range Manage. Agroforestry* 35 (1), 91–100.

Scotford, I., and Miller, P. (2004). Estimating tiller density and leaf area index of winter wheat using spectral reflectance and ultrasonic sensing techniques. *Biosyst. Eng.* 89 (4), 395–408.

Tao, H., Feng, H., Xu, L., Miao, M., Long, H., Yue, J., et al. (2020). Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors* 20 (5), 1296.

Thenkabail, P. S., Smith, R. B., and De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* 71 (2), 158–182.

Tilley, M. S., Heiniger, R. W., and Crozier, C. R. (2019). Tiller initiation and its effects on yield and yield components in winter wheat. *Agron. J.* 111 (3), 1323–1332.

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* 108 (50), 20260–20264.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., et al. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–a review. *ISPRS J. Photogrammetry Remote Sens.* 108, 273–290.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, G., et al. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and-3. *Remote Sens. Environ.* 118, 127–139.

Wang, W., Wang, X., Yao, C., Jin, T., Wu, J., and Su, W. (2020). Estimation of wheat planting density using UAV image. *Remote Sens. Land Resour.* 4), 111–119. doi: 10.6046/gtzyyg.2020.04.16

Wu, F., Wang, J., Zhou, Y., Song, X., Ju, C., Sun, C., et al. (2022). Estimation of winter wheat tiller number based on optimization of gradient vegetation characteristics. *Remote Sens.* 14 (6), 1338.

Wu, J.-H., Yue, S.-C., Hou, P., Meng, Q.-F., Cui, Z.-L., Li, F., et al. (2011). Monitoring winter wheat population dynamics using an active crop sensor. *Spectrosc. Spectral Anal.* 31 (2), 535–538. doi: 10.3964/j.issn.1000-0593(2011)02-0535-04

Xing, N., Huang, W., Ye, H., Dong, Y., Kong, W., Ren, Y., et al. (2021). Remote sensing retrieval of winter wheat leaf area index and canopy chlorophyll density at different growth stages. *Big Earth Data*, 1–23.

Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., and Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS J. Photogrammetry Remote Sens.* 150, 226–244.

Yu, R., Li, S., Zhang, B., and Zhang, H. (2021). A deep transfer learning method for estimating fractional vegetation cover of sentinel-2 multispectral images. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2021.3125429

Zenkl, R., Timofte, R., Kirchgessner, N., Roth, L., Hund, A., Van Gool, L., et al. (2021). Outdoor plant segmentation with deep learning for high-throughput field phenotyping on a diverse wheat dataset. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.774068

Zha, H., Miao, Y., Wang, T., Li, Y., Zhang, J., Sun, W., et al. (2020). Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning. *Remote Sens.* 12 (2), 215.

Zhang, B., Chen, Z., Peng, D., Benediktsson, J. A., Liu, B., Zou, L., et al. (2019). Remotely sensed big data: Evolution in model development for information extraction [point of view]. *Proc. IEEE* 107 (12), 2294–2301.

Zou, X., Zhu, S., and Mõttus, M. (2022). Estimation of canopy structure of field crops using sentinel-2 bands with vegetation indices and machine learning algorithms. *Remote Sens.* 14 (12), 2849.

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 09 January 2023
DOI 10.3389/fpls.2022.1075051

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service, United States Department of Agriculture (USDA), United States

REVIEWED BY
Guan Haiou,
Heilongjiang Bayi Agricultural University, China
Jianjun Du,
Beijing Research Center for Information Technology in Agriculture, China

*CORRESPONDENCE
Limin Shao
shaolimin@hebau.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 20 October 2022
ACCEPTED 19 December 2022
PUBLISHED 09 January 2023

CITATION
She L, Wang N, Xu Y, Wang G and Shao L (2023) Detection and counting of pigment glands in cotton leaves using improved U-Net.
Front. Plant Sci. 13:1075051.
doi: 10.3389/fpls.2022.1075051

COPYRIGHT
© 2023 She, Wang, Xu, Wang and Shao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Detection and counting of pigment glands in cotton leaves using improved U-Net

Lixuan She[1†], Nan Wang[1†], Yaxuan Xu[1], Guoning Wang[2] and Limin Shao[1*]

[1]State Key Laboratory of North China Crop Improvement and Regulation, College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, China, [2]State Key Laboratory of North China Crop Improvement and Regulation, College of Agronomy, Hebei Agricultural University, Baoding, China

Gossypol, as an important oil and raw material for feed, is mainly produced by cotton pigment gland, and has a wide range of applications in the fields of pharmaceutics, agriculture and industry. Accurate knowledge of the distribution of pigment gland in cotton leaves is important for estimating gossypol content. However, pigment glands are extremely small and densely distributed, manual counting is laborious and time-consuming, and difficult to count quickly and accurately. It is thus necessary to design a fast and accurate gland counting method. In this paper, the machine vision imaging technology is used to establish an image acquisition platform to obtain cotton leaf images, and a network structure is proposed based on deep learning, named as Interpolation-pooling net, to segment the pigment glands in the cotton leaf images. The network adopts the structure of first interpolation and then pooling, which is more conducive to the extraction of pigment gland features. The accuracy of segmentation of the model in cotton leaf image set is 96.7%, and the mIoU (Mean Intersection over Union), Recall, Precision and F1-score is 0.8181, 0.8004, 0.8004 and 0.8004 respectively. In addition, the number of pigment glands in cotton leaves of three different densities was measured. Compared with manual measurements, the square of the correlation coefficient (R^2) of the three density pigment glands reached 0.966, 0.942 and 0.91, respectively. The results show that the proposed semantic segmentation network based on deep learning has good performance in the detection and counting of cotton pigment glands, and has important value for evaluating the gossypol content of different cotton varieties. Compared with the traditional chemical reagent determination method, this method is safer and more economical.

KEYWORDS

cotton leaf, pigment gland, gossypol, U-Net, small objects, semantic segmentation

Introduction

Cotton is an important economic crop. It is not only the main source of high-quality natural fibers, but also the resource of protein and oil (Qu et al., 2017). Cotton pigment gland is a widely distributed unique tissue structure of cotton plants in most organs of cotton plants except pollen and seed coat. It is the dark brown opaque spots in cotton leaves (Liu et al., 2010; Zhang et al., 2021). Cotton pigment glands contain gossypol and other terpenoid aldehydes (Gao et al., 2020). Gossypol is a yellowish-brown polyphenol pigment insoluble in water but soluble in organic solvents, which is synthesized in the roots and carried in various organs of cotton plants (Zhao et al., 2020). Gossypol is widely used in pharmaceutical, agricultural and industrial application. For agricultural application, gossypol (Kong et al., 2010; Krempl et al., 2016) is used to control crop diseases and insect pests due to its good insect resistant characteristic and also for the control of rodent damage in the field due to its fertility resistance (Hahn et al., 1981). In pharmaceutics, the antifertile trait of gossypol is made use of in the manufacturing of contraceptives (Qian and Wang, 1984), drugs that inhibit the growth and proliferation of tumor cells (Badawy et al., 2007; Ni et al., 2013), and it has obvious effect on the treatment of gynecological diseases (Hsieh et al., 2020). However, the toxicity of gossypol limits its wide application as an important oil and feed raw material. Excessive intake of gossypol will cause human and animal digestive dysfunction and gastric mucosa damage (Lordelo et al., 2005; Gadelha et al., 2014). Therefore, the detection of gossypol content in cotton plants is very crucial in the improvement of the economic value of cotton.

At present, chemical reagent method is most commonly used for the detection of gossypol content such as High Performance Liquid Chromatography (HPLC) (Benbouza et al., 2002), Capillary Electrophoresis (Liu et al., 2005) and Ultraviolet Spectrophotometry (Przybylski et al., 2006). However, they will not only damage the samples, but also are costly. The relationship between cotton pigment glands and gossypol is close and complex. Pigment glands are the storage organ for gossypol and derivatives of gossypol. Their distribution density is positively correlated with the glanded cotton (Wilson and Smith, 1976). The number of pigment glands is an important predictor for estimating the gossypol content in glanded cotton. Pigment glands are small with an average diameter of 100 ~ 400 μm (Qian et al., 2017) and densely distributed, and the accurate counting of pigment glands is difficult to perform manually. At present, there are some methods for counting the pigment glands of cotton. The mainstream method used is the microscope observation method. Wang et al. (2018) used stereomicroscope to take pictures of the leaves of four species of cotton plants, *Gossypium barbadense L.*, *Gossypium hirsutum L.*, *Gossypium herbaceum L.*, and *Gossypium arboretum L.*, and counted the number of pigment glands in total, then divide it into the total area and got the number per square centimeter. This process is time-consuming and laborious, and the experience of labors may influence the result, and thus the accuracy of the count number can't be insured.

Computer vision or more precise plant phenotyping has been proved to be a reliable too in the field of plant biology, and the application of computer vision technology on the study of plant phenotype has achieved some fruitful results (Dee and French, 2015). The cotton pigment glands only account for a very small percentage in the whole cotton leaf image and so the small object detection method has to be used to detect the number of cotton pigment glands (Lin et al., 2014; Zhu et al., 2016). The coverage area that small objects in the image is generally dozens of pixels, or even a few pixels, with less information on the features and lacking feature expression ability. The small object detection method has always been the focal point of studies of many researchers (Li et al., 2022). At present, small object detection methods mainly are the traditional image processing method and deep learning method. The traditional image processing method mainly rely on manual analysis of image features, extraction of image features with algorithms, and distinguishing objects by feature values (Su et al., 2020). However, the feature extraction process is complex and time-consuming, and it is difficult to achieve the same detection accuracy as the detection of large object. With the improvement of computing power of equipment, deep learning has become more and more popular for image recognition and image segmentation (Tetila et al., 2019; Su et al., 2020). Different from the traditional image processing method, deep learning extracts image features from the network framework. Convolutional Neural Network (CNN) has excellent performance in exploring more deep information of images. Sun et al. (2021) introduced the segmentation masks to remove background images based on the original SSD model, which improved the detection performance of the Tsinghua-Tencent 100K and Caltech pedestrian dataset by adding context information. In Convolutional Neural Network, the shallow feature map has a small receptive field to detect small objects, while the deep feature map has a large receptive field to detect large objects. Zhou et al. (2018) proposed a novel Scale-Transferable Detection Network (STDN) by embedding a super-resolution layer in a Dense Convolutional Network (DenseNet). Firstly, the small-scale feature map is obtained, and then the large-scale feature map is obtained by reducing the number of channels of the feature map in the super-resolution layer to explicitly explore the inter-scale consistency. However, the objects are closely adjacent to each other in the single image where the small objects have dozens or even hundreds of objects to be detected (Ming, 2021), which is easy to cause detection overlapping and false detection of small objects (Goldman et al., 2019; Pan et al., 2020), not to mention the identification and counting of dense small objects, which is more challenging.

Pigment glands are small, numerous and densely distributed, so it is difficult to manually count them accurately. On the basis of previous research work, we inverted the structure of the

classical U-Net network and designed a semantic segmentation model of pigment glands based on deep learning to realize the rapid recognition and counting of densely distributed pigment glands. The model first performs the interpolation operation for up-sampling to amplify the image features, and then performs the pooling operations to achieve multi-scale fusion of the network by fusing deep semantic information and shallow representation information. First of all, a cotton leaf image acquisition device was built to collect the RGB image of cotton leaf with a color industrial camera, and then the image was manually annotated, binary image conversion, rotation and cropping, so as to build the data set of cotton leaves. After that, the improved model was trained and validated. The shape feature filtering method was used to optimize the segmentation results to reduce the error caused by the leaf vein shadow in the image. Finally, the pigment glands in the images were counted. The flow chart is shown in Figure 1. In order to evaluate the segmentation performance of the network, we compared the segmentation and counting results of the improved model with original U-Net, DeepLabv3+ and manual counting. The comparison results showed that the improved model had achieved high-throughput and accurate detection of cotton pigment glands.

Materials and methods

Image collection

In August 2021, cotton leaf image acquisition experiments were carried out at the innovation experimental park of Hebei Agricultural University in Baoding City (38.85°N, 115.30°E), Hebei Province, China. The true leaves of seedling stage of upland cotton with a growth cycle of 35 days were taken as samples, among which those with intact leaf phenotype and uniform chlorophyll distribution were selected for cleaning and drying.

An image acquisition platform was established in advance, as shown in Figure 2. The cotton leaf was placed under the white square light and the color mesh industrial camera was connected (MV-GED200C-T, Mind Vision, China, maximum resolution: 1600×1200) to the computer. The camera support knob was adjusted to keep the camera at an appropriate height so that the leaf can be in the camera aperture completely, and extend in as much as possible to collect cotton leaf images of different densities. The focal length and aperture of the camera lens (OPT, focal length: 12mm) were adjusted, as well as the light source controller to make the pigment glands in the cotton leaf as clear as possible. Finally, the RGB images of cotton leaves were collected by the industrial camera controlled by the industrial camera software.

Data set processing

33 RGB images of cotton leaves with a resolution of 1600×1200 were collected by the image acquisition device, and saved as BMP file. The collected 33 cotton leaf images were screened to eliminate the ones that were blurred, damaged and dark leaf images, and 9 qualified cotton leaf images were sifted out and labeled manually. The image annotation was performed by annotation tool LabelMe (Russell et al., 2008), it contained about 37,584 labeled pigment glands in 9 images, and the average labeling time of each image was about 5.5 hours. The

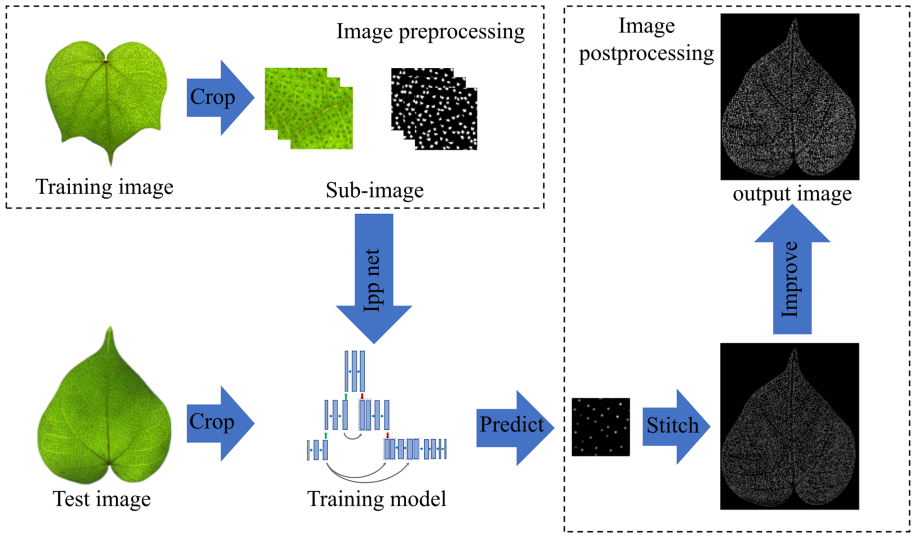


FIGURE 1
Flow chart of cotton pigment gland detection.

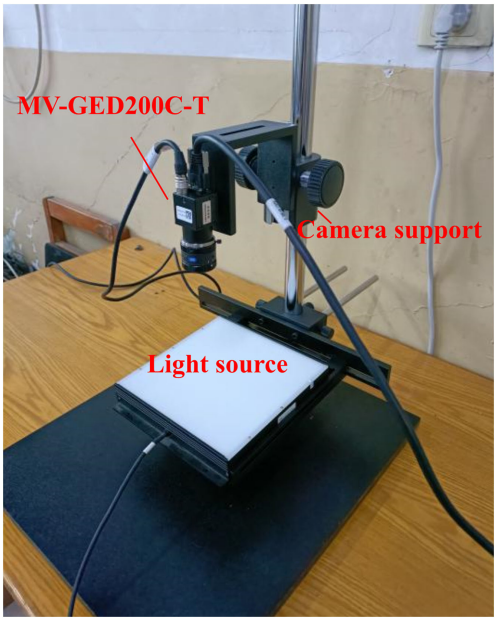


FIGURE 2
Cotton leaves image acquisition device.

json files generated after labeling were saved in the computer. In order to improve the generalization performance of the training model, the 9 labeled json files were converted into binary images of black and white, where the pigment glands were white and the background was black, as shown in Figure 3. Then, the binary images were reversed horizontally, vertically and mirrored, and 36 images were generated. Because the leaf image field of view was extensive, to allocate GPU memory was difficult. To solve the problem, the image was cropped into smaller size and 6912 sub-images of 100×100 of pixel size were generated. 80% of them were used as the training set and 20% as the validation set.

In addition, we also selected another 7 clear images of different densities as the counting study of this model, and the 7 images were cropped into 453 sub-images. In order to verify the counting accuracy of pigment glands with different densities, the sub-images were divided into three grades according to the density. It is assumed that the number of pigment glands in 4.5×4.5mm area was less than 40 be level 1, the number more than 40 but less than 80 be level 2, and the number more than 80 be level 3. The three levels were categorized into three different datasets DS1, DS2 and DS3 respectively, and the number of sub-images of DS1, DS2 and DS3 was 184, 156 and 113 respectively. The sub-images of three datasets with different densities are shown in Figure 4.

Model improvement

In the image, the pigment glands occupy a small area, so the object detection method is more challenging in locating the bounding box than the large and mesoscale objects. In the prediction process, if the prediction bounding box is shifted by one pixel, the impact on the small object will be much higher than that on the large and mesoscale object, so it is difficult to achieve high detection accuracy. The method of image segmentation is to classify each pixel in the image at pixel level. So, we used a semantic segmentation model based on deep learning to classify the leave images into pigment glands and non-pigment glands at pixel level. Compared with the manual object detection method, the label map generated by semantic segmentation was simpler, and thus easier to stitch the sub-images together, and the context information was enhanced better.

U-Net was originally used in medical image segmentation. Because the network structure fused the deep features and shallow features with skip connections, U-Net is more effective in dealing with the complex segmentation. U-Net is a process of encoding and decoding, mainly composed of contracting path, expansive path and prediction network (Ronneberger et al., 2015). The contracting path is composed of five effective feature layers, and each of them performs two 3×3

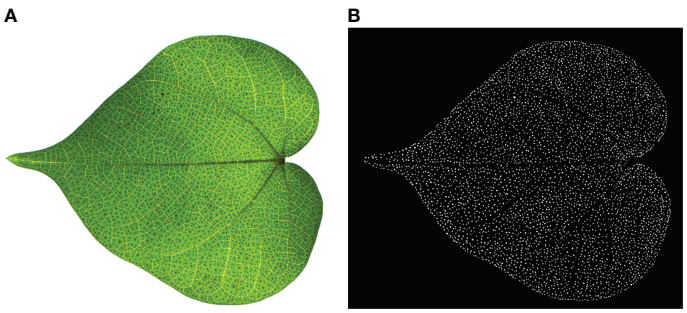


FIGURE 3
Manual labeling of cotton leaves at seedling stage. **(A)** Original image; **(B)** Annotated image. The background is in black, and the cotton pigment glands are in white.

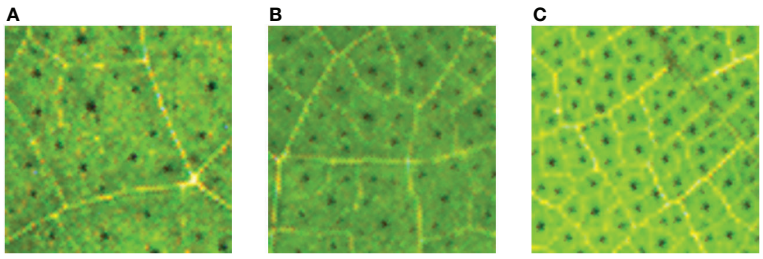


FIGURE 4
The sub-images of three datasets with different densities. **(A)** DS1; **(B)** DS2; **(C)** DS3.

convolution operations and one 2×2 maximum pooling operation for feature extraction. The structure of extraction is to reduce the feature images. After a series of convolution and pooling operations, the size of the feature image was halved and the number of channels increased. The expansive path is symmetric with that of the contracting path, and each effective feature layer was subjected to two 3×3 convolutions and one 2×2 deconvolution operation, that is to say, the five effective feature layers obtained from the feature extraction network were for up-sampling to magnify the reduced image. Finally, the corresponding effective feature layers in the contracting path and the expansive path were spliced and fused through skip connections. The fused feature layer combines the complex information extracted from the deep network and some simple information from the shallow network, such as edge feature, so that the network can handle more complex segmentation tasks. Then the final feature layer fused during feature extraction and enhanced feature extraction was predicted, and the input images were classified on the pixel level. During prediction, the number of channels was converted once, resulting in the number of output channels to be the number of categories classified.

In this experiment, the cotton pigment glands were the only a few pixels in the leaf image. The cotton leaf image will be downscaled by half every time while U-Net model was down-sampled, making feature extraction more difficult, and the feature loss more serious during the training of the model and thus impair the successful realization of cotton pigment glands segmentation. We inverted the classic U-Net structure with a conical structure of interpolation followed by pooling instead of the original "U" structure of down-sampling followed by up-sampling, and the improved network model was named interpolation-pooling net (abbreviated as Ipp Net). This network structure can effectively suppress the loss of pigment gland features caused by the pooling operations. The improved network model has three chief components, up-sampling, down-sampling and network prediction. The Ipp Net model structure is shown in Figure 5.

Up-sampling. The sub-images of cotton leaves were entered into the model as input. After two 3×3 convolution layers and batch standardization, the nonlinear ability was strengthened through the ReLU activation function. Then the nearest-neighbor interpolation method was used to amplify the image features. The nearest pixel among the 4 pixels around the pixel to be interpolated was selected as the target pixel and inserted into the amplified image, and the final image was amplified from the original 120×120 to 240×240, and the number of channels increased to 16. After convolutions and interpolation operation, the feature image size was doubled again to 480×480, and the number of channels remained unchanged. For each convolution and interpolation operation, the image size doubled once. The un-sampling process of feature extraction was to enlarge the feature map. Unlike the traditional U-Net model, the improved Ipp Net adopts the Same convolution. The size of the convolution kernel won't affect the size of the feature map after convolution, but not the step size used for convolution, which was set as 1. The result shows that the size of the output feature map after convolution was the same as the input image. In this mode, the size of the feature image is unchanged during forward propagation, and the convolution operation does not need to accurately calculate the size of the image.

Down-sampling. The feature extraction after the feature images were amplified was performed using convolution and max pooling operations. The pooling method used max pooling to retain the salient features of the image and reduce the feature dimension, so that the model can learn the edge and texture structure of pigment glands, and stably segment the phenotypic traits of pigment glands. After two convolutions and one pooling, the salient features of the image were retained, and the image size was reduced from 480×480 to 240×240. Then, two more convolutions and one pooling operation are performed, and finally reduced to the size of the input image. After each set of convolution and pooling operations, the size of the feature image was reduced to half of the original size. The improved network is also a cross-layer connection, the Concat dimension splicing and fusion method are adopted to fuse the primary features corresponding to the up-sampling and the down-sampling, and the channel number was also increased the same as the corresponding feature layer. On the one hand, the fused feature layer recovered the lost image position information during down-sampling, on the other hand, the low-level details

10.3389/fpls.2022.1075051

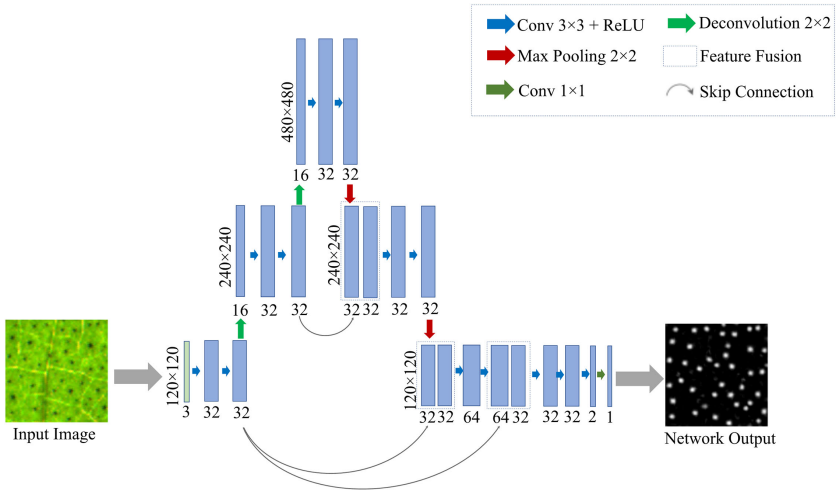


FIGURE 5
Ipp Net model. The structure of up-sampling and down-sampling is adopted, and the up-sampling and down-sampling are performed twice respectively to obtain image features.

from different scale feature maps were combined with high-level semantics, which enriched the feature information.

Network prediction. The last layer of the network is a 1×1 convolution layer. The feature vector can be converted into the number of required classification results in the layer. The results in the output layer of the network are transformed into nonlinear values by ReLU activation function, that is, the values less than 0 become 0, and the values greater than 0 are assigned as 1, which is used for the conversion of discrete probability values to binary. Finally, the output is single-channel.

Image cropping and stitching

Since the leaf image contains a wide-angle view, which makes GPU memory allocation difficult, and also because that the global semantic information of the pigment gland is not so important in the pigment gland segmentation, even a portion of an intact leaf is taken out, the pigment gland can still be identified. Therefore, we cropped each image into sub-images with the size of 100×100, as shown in Figure 6, and a total of 6912 sub-images were cropped. Image cropping reduced the GPU pressure on the one hand and expanded the data set on the other hand. However, in the process of stitching, we found that the undetectable white stripe will appear on the edge. To identify it, we expanded the original sub-image of 100×100 to 120×120 and fed into the network. When the part of 4×5 pixels on the sub-image edge was cropped off, and the remaining images were stitched, which effectively suppressed the impact of edge defects on segmentation, and reduced the counting error caused by incomplete pigment glands at the cropped edges.

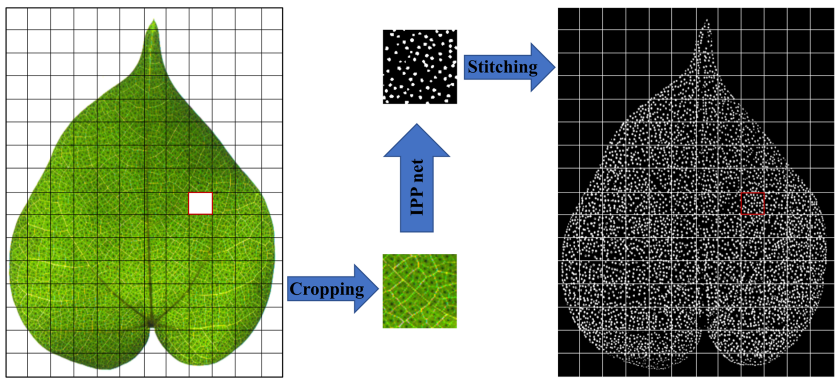


FIGURE 6
Image cropping and stitching. The red box was one of the sub-images, which has been cropped and segmented and then stitched.

Result

Model parameter

Python 3.6 was used as the server environment for network training to train and test the model under TensorFlow 1.14.0 and CUDA 10.0. The server was equipped with NVIDIA GeForce GTX 1080Ti graphics processing unit for acceleration and 16GB video memory.

Mean square error (MSE) is used as the loss function of the model to evaluate the difference between the predicted value and the real value of the model, so as to provide the model with an object that can be optimized and make the optimizer move in the right direction.

$$\mathrm{MSELoss} = \frac{1}{n}\sum_{k}(y_k - t_k)^2 \tag{1}$$

where y_k s the output of the network, t_k the label value of training data, and k represents the size of the training set, n is the number of samples.

To make each parameter of the loss function reach the most appropriate value and make the loss function as smooth as possible, we used the adaptive motion estimation (Adam) optimizer to dynamically adjust the learning rate of each parameter using the first-order moment estimation and second-order moment estimation of the gradient. Under Adam, the learning rate bias was corrected, and the parameters were more stable and better adapted to the problem of gradient sparsity. The parameter settings of training model are shown in Table 1.

Evaluation indicators

The segmentation results of the model are shown in Figure 7. It is known from Figure 7 that the model has a good effect on the identification of pigment glands in cotton leaves. In order to objectively and reasonably evaluate the effect of the network model in the pigment gland segmentation of cotton leaves, four model evaluation indicators were introduced to evaluate the segmentation effect of the model and they were mIoU (Mean Intersection over Union), Precision, Recall and F1-score respectively. Four evaluation indicators defined by the equations from (2) to (5) were calculated by the confusion matrix, as shown in Table 2, the columns in the table represent the predicted values of the model, and the rows represent the label values of the model.

$$\mathrm{mIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{2}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{3}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{4}$$

$$\mathrm{F1}-score = \frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} = \frac{2\times\mathrm{TP}}{2\times\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{5}$$

Where, True Positive (TP) is the number of the correctly identified pigment gland pixels defined by the model; False Negative (FN) refers to the number of pixels wrongly identified by the model as the background; False Positive (FP) indicates the number of pixels incorrectly identified as the pigment gland, True Negative (TN) is the number of pixels of the background correctly identified as the background.

Mean Intersection over Union (mIoU) is a common evaluation metric for semantic image segmentation, which first computes the IoU for each semantic class and then computes the average over classes. It is a commonly used metric for the measurement of the image segmentation performance of the algorithm at pixel level. It is also used to compare the similarities and differences between the segmentation results and the label set. The Precision indicates that the proportion of pigment glands predicted by the model is close to the actual result. Recall indicates how many positive examples in the sample are predicted correctly. F1-score, also called balanced F score, is defined as the harmonic mean of Precision and Recall. During the model training, mIoU, Recall, Precision and F1-score of each epoch output is calculated in detail using the validation set, followed by the model performance evaluation using the test set not involved in the training.

TABLE 1 The parameter settings of training model.

Parameter	Value
optimizer	adam
Epochs	60
Steps per epoch	250
Batch size	2

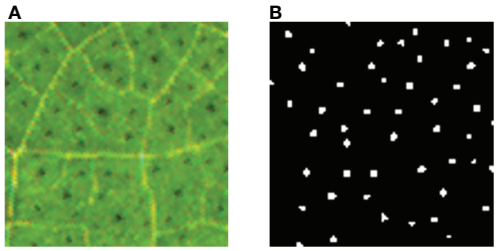


FIGURE 7
Segmentation results. **(A)** Original image; **(B)** Segmented result.

TABLE 2 Confusion matrix.

P\L	Positive	Negative
Positive	42681	10644
Negative	10645	666030

After each epoch, the accuracy and loss values are calculated in the training set and validation set to monitor the fitting degree of the model. The total model training last about 9.5 h. The loss value of the model appeared to be flat at the 40th epoch and the accuracy of the model stabilized at 0.967; the loss value was finally stabilized at 0.0238, mIoU, Recall, Precision and F1-score was 0.8181, 0.8004, 0.8004 and 0.8004 respectively.

In addition, two other classical models, U-Net and DeepLabv3 +, were trained and compared with the Ipp Net in this paper. The segmentation results are shown in Figure 8. The experiment results based on the performance measurement evaluation are shown in Table 3. Figure 9 is the scores of box plot of the three models in the different epochs of the test set. It mainly contains six data nodes, which respectively calculate the upper edge upper quartile Q3, median, lower quartile Q1, lower edge and outliers of test sets in different epochs of mIoU, Recall, Precision and F1-score. Compared with the original U-Net, mIoU, Recall, Precision and F1-score increased by 9.51%, 9.51%, 21.78% and 16.24% respectively, and which was 14.78%, 7.98%, 37.43% and 26.49% better than DeepLabv3+. The segmentation results show that U-Net and DeepLabv3+ are not effective in the segmentation of cotton pigment glands with small objects and dense distribution. In contrast of the model used in this paper, U-Net and DeepLabv3+ will cause the feature loss of pigment glands during feature extraction, so the lost feature information will be classified as background during pixel classification. The model in this paper adopts the structure of interpolation and pooling, which is to amplify the features of the image first and then extract the image features. Therefore, the Ipp Net model has a better segmentation effect in dealing with small objects.

Results optimization

Due to the vein shadows in cotton leaves which are similar to pigment glands, they will be classified as foregrounds, thereby reducing the accuracy of the model. At this time, if the pigment glands are directly counted from the segmentation results, some leaf vein shadows will also be inclusive and counted, and the counting result will be much higher than the actual number of pigment glands. By optimizing the segmentation results, the mistakenly inclusive vein shadows can be filtered, thus weaken the influence of vein shadows on the final counting results. The segmented leaf images were optimized using the machine vision software HALCON 17.12.0.0, and the optimization results are shown in Figure 10. Select_shape is an operator that can select the regions according to the shape features such as the area and roundness of the input connected domain. Its usage is simple and has good effect on image filtering and optimization. The diameter of the pigment glands is generally 100~400μm, approximately a few pixels as an image. 3 pixels was the minimum pixel value of the pigment glands in the sample image and we optimized the image with shape feature selection and morphological filtering. The connected domains with area parameters between 2-938 744 pixels were selected and counted, while the connected domains with area parameters less than 2 pixels were filtered out. Then, the count_obj operator was used to count the connected domains selected by the select_shape operator, and the optimized image was saved to count the pigment glands.

The image was optimized by the combination of morphological filtering and shape feature screening, which eliminated the shadow impurities of leaf veins, and at the same time preserved the

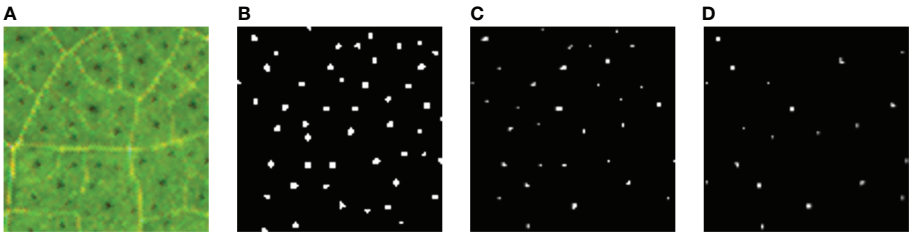


FIGURE 8
Comparison of segmentation results of the three models. **(A)** Original image; **(B)** The segmentation result of Ipp Net; **(C)** The segmentation result of U-Net; **(D)** The segmentation result of DeepLabv3+.

TABLE 3 Evaluation indicators.

Model	Acc	mIoU	Precision	Recall	F1-score
U-Net	0.995	0.723	0.7053	0.5826	0.638
DeepLabv3+	0.994	0.6703	0.7206	0.4261	0.5355
Ipp Net	**0.967**	**0.8181**	**0.8004**	**0.8004**	**0.8004**

The bold values are the improved algorithm and corresponding evaluation indicators.

characteristics of the pigment glands in the image. Compared with other filtering methods, such as median filtering, shape feature screening, this method can well preserve the features of very small pigment glands.

Counting results

In order to better evaluate the Ipp Net model, the pigment glands of 453 sub-images in three datasets DS1, DS2 and DS3 used for model evaluation were segmented and counted, Figure 11 shows the comparison results of Ipp Net, U-Net and DeepLabv3+ with manual counting at three densities. The abscissa in Figure 11 is the result of manual counting, and the ordinate is the counting result of Ipp Net, U-Net and DeepLabv3 +. As shown in Figure 11, the square of the correlation coefficient (R^2) of Ipp Net reached 0.97, 0.94 and 0.91 respectively in the three datasets, DS1, DS2 and DS3, which were 0.03, 0.11, 0.21 and 0.04, 0.06, 0.28 higher than those of U-Net and DeepLabv3+, respectively. With the increase of pigment gland density, the detection and counting ability of the three models decreased, but Ipp Net still performed well in the detection and counting high-density pigment glands, and R^2 still reached 0.91. The results show that Ipp Net has the best fitting degree and is more accurate than U-Net and DeepLabv3+ at different densities. The proposed method can realize high-throughput detection and counting of cotton pigment glands.

Discussion

Gossypol is a peculiar substance of Gossypium plants. It is toxic and has extremely high research value in agriculture, pharmaceutics and other fields. At present, the detection of

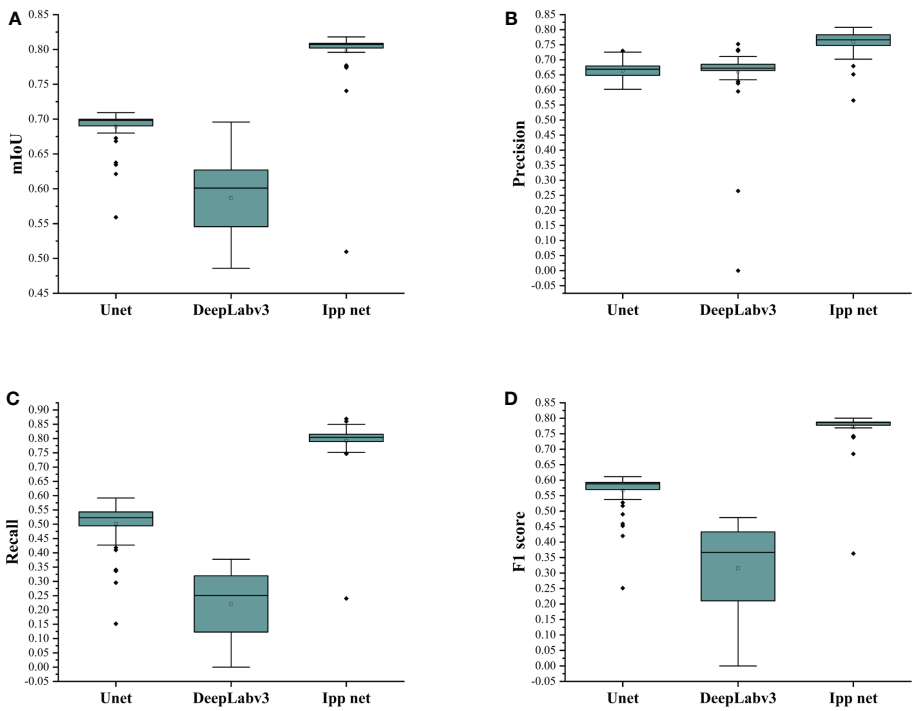


FIGURE 9
Evaluation indicators of U-Net, DeepLabv3+ and Ipp Net models of different epochs. **(A)** mIoU; **(B)** Precision; **(C)** Recall; **(D)** F1-score.

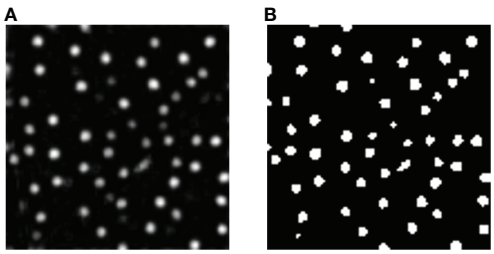


FIGURE 10
Segmentation results optimization. **(A)** Segmentation results of Ipp Net; **(B)** Optimized images of the segmentation results.

gossypol mainly leverages High Performance Liquid Chromatography (HPLC), spectrophotometry, Capillary Electrophoresis and other chemical reagent methods. These methods require to break up the cotton organ samples for detection. Although the result can achieve high accuracy, they are not economical and non-destructive. Plant phenotypic analysis has made important progress in crop identification and detection. Pigment gland is the main carrier of gossypol. There is a significant correlation between the phenotypic traits of pigment gland and gossypol content, which can be used as a basis for measuring the amount of gossypol in phenolic gossypol. Pigment glands are small and densely distributed. At present, people mostly analyze the distribution of pigment glands by hand, and the workload is huge. To reduce work, we proposed a machine vision method to detect the number of pigment glands in cotton leaves. Compared with the manual method, it is more convenient and detect faster, and the distribution of pigment glands in cotton leaves can be analyzed more accurately.

The densely distributed cotton pigment gland is only a few pixels in the image of cotton leaves, and the detection is counted as small object detection. Due to the small coverage area in the image and hardly available features of small objects that are distributed in a dense manner, it is difficult to locate the general large objects. At the same time, the predicted bounding box may also filter out a large number of correct ones due to the non-maximum suppression operation during post-processing,

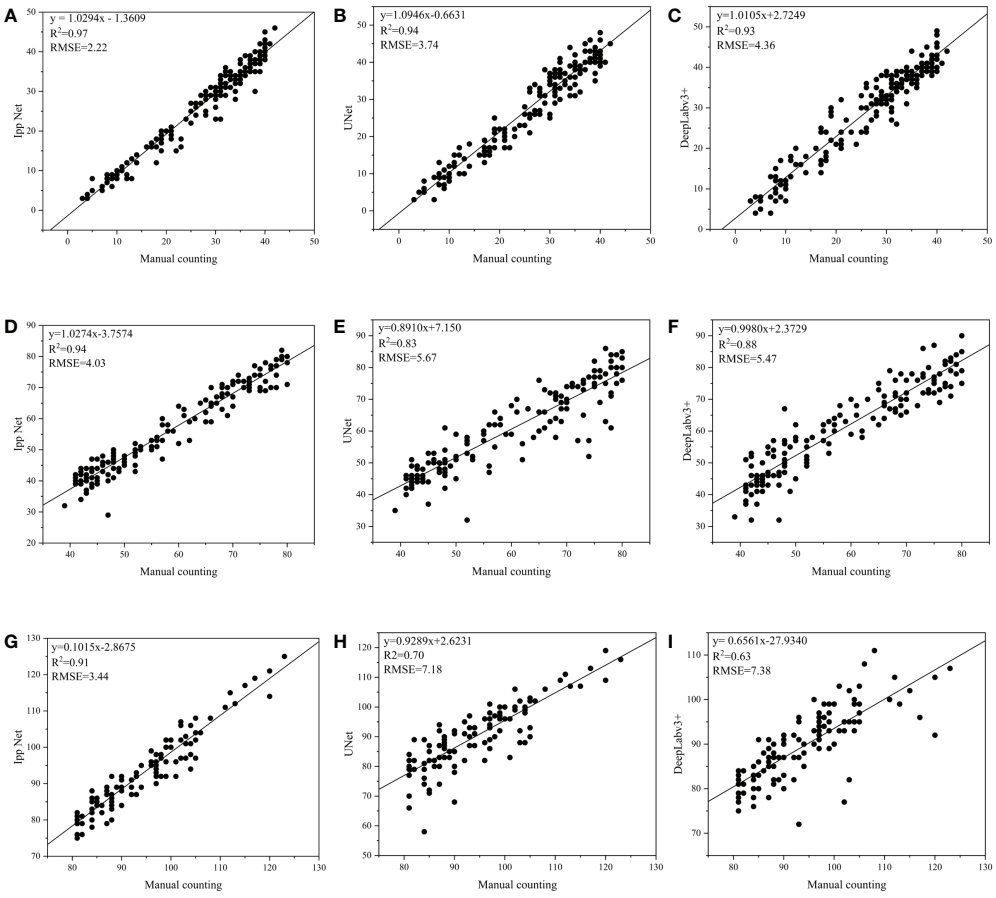


FIGURE 11
Comparison results of Ipp Net, U-Net and DeepLabv3+ with manual counting at three different densities. **(A–C)**: DS1, low-density; **(D–F)**: DS2, medium-density; **(G–I)**: DS3, high-density.

resulting in overlooking detection. Therefore, some commonly used convolutional neural networks cannot achieve good detection results for the detection of dense small objects. The U-Net network performs feature fusion of different dimensions in the channel dimension for the network to segment and detect images of different sizes. In the middle of detection, the object will be aggregated at one point in the feature image after multiple times of down-samplings, which makes the feature extraction more difficult. Therefore, the detection effect on the cotton leaf pigment gland using the U-Net network is not accurate.

In this paper, a semantic segmentation model based on interpolation-pooling network is proposed to materialize automatic segmentation of cotton pigment glands, and count the segmentation results. The improved Ipp Net is in the sequence of interpolation first and then pooling to avoid the target feature information loss caused by down-sampling in the classic U-Net network structure, and improve the detection accuracy of the network for small objects. The experiment results show that the improved Ipp Net has high segmentation accuracy of 0.967 for the cotton leaf image data set. In order to verify the ability of Ipp Net to detect the pigment glands with different densities, three data sets were made according to the density distribution of pigment glands in cotton leaf sub-images to test the counting results of the model. Compared with the manual counting methods, the detection accuracy of Ipp Net is slightly lower than that of manual counting, but the required detection time is shorter and labor is saved. For the trained model, the average detection time of each sub-image is only 202 ms, and it takes about 91.5 s to detect the 453 sub-images, while the manual method takes about 11 hours to complete the statistical task of pigment glands. Therefore, the semantic segmentation model proposed in this paper can replace the manual method for counting the cotton pigment glands.

Different from other small object detection methods, we used the semantic segmentation method to detect the pigment glands, and used the object detection method to detect small objects with very dense distribution. After sampling for several times, the small objects adjacent to the aggregation area would be aggregated together in the deep feature map, resulting in the model being difficult to distinguish. The boundary distance between small objects in the aggregation area is too close, which will lead to the difficulty of bounding box regression and the difficulty of model convergence. It is difficult to achieve the ideal detection effect. The method of semantic segmentation is to classify each pixel in the image, so that the detection of pigment glands is more accurate, and the counting result of pigment glands is closer to the real value.

Before constructing the Ipp Net model, we used the traditional image segmentation methods such as threshold segmentation to detect the pigment glands in cotton leaves. However, due to the small size and dense distribution, the traditional image segmentation method was not ideal for the segmentation of pigment glands. Especially in some images containing veins and leaf edges, the detection results were too different from the real values, and the pigment gland counting task cannot be completed at all. In addition, the pigment glands in the image were basically unrecognized in the intact leaves with a wide field of view.

However, the model still suffers from false detection of pigment glands. The convolutional neural network model is easily affected by the quality of image annotation. The main reason for that is because of the veins and leaf edges in cotton leaves. When labelling the pigment glands on the leaf veins, some of the veins will also be annotated. In this way, the model will also learn from wrong labelling. Affected by the white light, the edges of the leaf were too bright, making it more difficult to mark manually, so the leaf edge will also cause false detection. In addition, the color of the pigment gland was very close to vein shadow, making it even harder for the computer to distinguish. Consequently, some vein shadows are mistakenly identified as pigment glands in the model, thus reducing the accuracy of the model. When the semantic segmentation of general objects is carried out in neural network, the edges of the object in the image have been through pixel mutation and contain richer feature information than the inside of the object, so it is easier to recognize them. The pigment glands have fuzzy edges which contain very little information. Although the pigment glands in the image are easy to identify, the edge segmentation is usually not ideal. On the other hand, the closing operation of morphological filtering is used for image optimization, which leads to some adjacent pigment glands forming a connected domain of adhesion after closing operation. As shown in Figure 12, A is the leaf image segmented by the improved Ipp Net, and B is the optimized image, and the counting result will be smaller than the true value.

Image enhancement of the data set is an important method to improve the accuracy of convolutional neural network. Illumination and noise in the imaging process are important influential factors on the detection accuracy of the model. Although the IPP Net model has a good effect in identifying

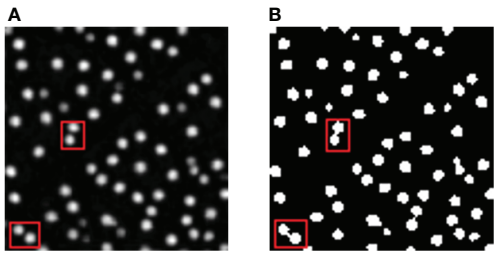


FIGURE 12
Cases of misjudgment. **(A)** The segmentation result of Ipp Net model; **(B)** The result after filtering and optimization. The red box is the part of the image optimized to cause the adhesion of pigment glands.

the pigment glands in cotton leaves, its segmentation effect is easily affected by the veins shadow of cotton leaves and the image noise of leaf edge. The results were optimized by shape feature filtering, but there is still some gap between the automatic counting results and the true value. In the follow-up research, on the one hand, the cotton leaf data set was smoothed and filtered to eliminate the influence of veins and their shadows on the model accuracy as much as possible while retaining the characteristics of pigment glands; on the other hand, we will proceed to further improve the model and try to integrate the global context information into the detection model, so as to improve the detection accuracy of the model in the veins to minimize the influence of vein shadow. In addition, the function to calculate the area of pigment glands and to estimate the content of Gossypol in cotton leaves by combining the area and quantity will be added to the model. A non-destructive and accurate estimation model of Gossypol content will be constructed, and then the model of cotton disease resistance, insect resistance, gland phenotypic traits and physical and chemical parameters of Gossypol content was established to realize the auxiliary decision-making technology of disease resistance and insect resistance quality identification.

Conclusions

Cotton pigment gland is the main carrier of gossypol, an important structure for studying gossypol and its derivatives. Its density and size reflect the amount of gossypol in glanded cotton. Cotton pigment glands are extremely small and densely distributed. At present, researchers mostly use microscope observation to estimate the density of pigment glands based on artificial experience. The manual counting method is easily influenced by the experience of researchers and is time-consuming and laborious, which brings great inconvenience to the study of cotton pigment glands and gossypol.

In this paper, a neural network model for automatic detection and counting of pigment glands in cotton leaves is proposed, aiming to detect small and densely distributed pigment glands in cotton leaves by computing of the semantic segmentation model. The interpolation-pooling network is proposed, and the practice of interpolation first and then pooling can effectively avoid the loss of target feature information caused by convolution, which is more conducive to the extraction of small objects information with small individual or small proportion in the image. The model was validated using the images of glanded cotton true leaves, and mIoU, Recall, Precision and F1-score were used to evaluate the network performance. The final validation set scores were 0.8181, 0.8004, 0.8004 and 0.8004 respectively. Compared with U-Net and DeepLabv3+, the mIoU, Recall, Precision and F1-score of Ipp Net are higher by 9.51%, 9.51%, 21.78%, 16.24% and 14.78%, 7.98%, 37.43% and 26.49% respectively. The segmentation has better outcome. In addition, the detection ability of the Ipp Net model in cotton pigment glands of different densities was analyzed, and found that it has good performance in different densities. The results show that the Ipp Net model has better segmentation effect and good robustness, and can replace manual counting for cotton pigment gland counting in some researches, which provides an important practical reference for the study of cotton pigment gland.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

LShe, NW and LShao conceived the idea and proposed the method. LShe, YX, and GW contributed to the preparation of equipment and acquisition of data. LShe, YX, and NW wrote the code and tested the method. LShe and YX validated the results. LShe wrote the manuscript. LShe, NW and LShao revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was funded by State Key Laboratory of North China Crop Improvement and Regulation (NCCIR2022ZZ-15, NCCIR2021ZZ-23), Science and Technology Project of Hebei Education Department (BJ2021058) and Basic Scientific Research Business Fee Project of Hebei Provincial Colleges and Universities (KY2021037).

Acknowledgments

We would like to thank KetengEdit (www.ketengedit.com) for its linguistic assistance during the preparation of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1075051/full#supplementary-material

References

Badawy, S. Z. A., Souid, A. K., Cuenca, V., Montalto, N., and Shue, F. (2007). Gossypol inhibits proliferation of endometrioma cells in culture. *Asian J. Androl* 9, 388–393. doi: 10.1111/j.1745-7262.2007.00168.x

Benbouza, H., Lognay, G., Palm, R., Baudoin, J. P., and Mergeai, G. (2002). Development of a visual method to quantify the gossypol content in cotton seeds. *Crop Sci.* 42, 1937–1942. doi: 10.2135/cropsci2002.1937

Dee, H., and French, A. (2015). From image processing to computer vision: plant imaging grows up foreword. *Funct. Plant Biol.* 42, III–IIV. doi: 10.1071/FPv42n5_FO

Gadelha, I. C. N., Fonseca, N. B. S., Oloris, S. C. S., Melo, M. M., and Soto-Blanco, B. (2014). Gossypol toxicity from cottonseed products. *Sci. World J.* 2014, 231635. doi: 10.1155/2014/231635

Gao, W., Xu, F. C., Long, L., Li, Y., Zhang, J. L., Chong, L., et al. (2020). The gland localized CGP1 controls gland pigmentation and gossypol accumulation in cotton. *Plant Biotechnol J* 18, 1573–1584. doi: 10.1111/pbi.13323

Goldman, E., Herzig, R., Eisenschtat, A., Goldberger, J., and Hassner, T. (2019). "Precise detection in densely packed scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (California: IEEE). arXiv [Preprint] Available at: https://arxiv.org/abs/1904.00853.

Hahn, D. W., Rusticus, C., Probst, A., Homm, R., and Johnson, A. N. (1981). Antifertility and endocrine activities of gossypol in rodents. *Contraception* 24, 97–105. doi: 10.1016/0010-7824(81)90072-X

Hsieh, Y. S., Chu, S. C., Huang, S. C., Kao, S. H., and Chen, P. N. (2020). Gossypol reduces metastasis and epithelial-mesenchymal transition by targeting protease in human cervical cancer. *Am. J. Chin. Med.* 49, 181–198. doi: 10.1142/S0192415X21500105

Kong, G. C., Daud, M. K., and Zhu, S. J. (2010). Effects of pigment glands and gossypol on growth, development and insecticide-resistance of cotton bollworm (Heliothis armigera (Hubner)). *Crop Prot* 29, 813–819. doi: 10.1016/j.cropro.2010.03.016

Krempl, C., Heidel-Fischer, H. M., Jimenez-Aleman, G. H., Reichelt, M., Menezes, R. C., Boland, W., et al. (2016). Gossypol toxicity and detoxification in helicoverpa armigera and heliothis virescens. *Insect Biochem. Molec Biol.* 78, 69–77. doi: 10.1016/j.ibmb.2016.09.003

Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., et al. (2014) *Microsoft COCO: Common objects in context. 2014 European conference on computer vision*. Available at: https://arxiv.org/abs/1405.0312v3.

Liu, E. B., Zhang, M., Xue, B. C., Zhao, P. X., and LIU, Y. (2005). Detection of ultra trace amount gossypol with chemiluminescence using capillary electrophoresis as injection techniques. *Chin. Chem. Lett.* 16, 1649–1651. doi: 10.1201/b19680-10

Liu, W. Z., Zhou, Y. F., Wang, X., and Jiao, Z. J. (2010). Programmed cell death during pigment gland formation in gossypium hirsutum leaves. *Plant Biol.* 2010 (12), 895–902. doi: 10.1111/j.1438-8677.2009.00291.x

Li, K. C., Wang, X. Q., Lin, H., Li, L. X., Yang, Y. Y., Meng, C., et al. (2022). Survey of one-stage small object detection methods in deep learning. *Technol.* 16, 41–58. doi: 10.3778/j.issn.1673-9418.2110003

Lordelo, M. M., Davis, A. J., Calhoun, M. C., Dowd, M. K., and Dale, N. M. (2005). Relative toxicity of gossypol enantiomers in broilers. *Poult Sci.* 84, 1376–1382. doi: 10.1093/ps/84.9.1376

Ming, H. (2021). *Research and application of small object detection based on deep learning* ([Mianyang]: Southwest University of Science and Technology: [master's thesis]). doi: 10.27415/d.cnki.gxngc.2021.000812

Ni, Z. H., Dai, X. F., Wang, B., Ding, W., Cheng, P. K., Xu, L., et al. (2013). Natural bcl-2 inhibitor (-)-gossypol induces protective autophagy *via* reactive oxygen species-high mobility group box 1 pathway in burkitt lymphoma. *Leuk Lymphoma* 54, 2263–2268. doi: 10.3109/10428194.2013.775437

Pan, X. J., Ren, Y. Q., Sheng, K. K., Dong, W. M., Yuan, H. L., Guo, X. W., et al. (2020) *Dynamic refinement network for oriented and densely packed object detection. arXiv [Preprint]*. Available at: https://arxiv.org/abs/2005.09973v2 (Accessed June 10, 2020).

Przybylski, P., Bejcar, G., Huczynski, A., Schroeder, G., Brzezinski, B., and Bartl, F. (2006). 1H- and 13C-NMR, FTIR, UV-VIS, ESI-MS, and PM5 studies as well as emission properties of a new schiff base of gossypol with 5-methoxytryptamine and a new hydrazone of gossypol with dansylhydrazine. *Biopolymers* 82, 521–535. doi: 10.1002/bip.20505

Qian, S. Z., and Wang, Z. G. (1984). Gossypol: A potential antifertility agent for males. *Ann. Rev. Pharmacol. Toxicol.* 24, 329–360. doi: 10.1146/annurev.pa.24.040184.001553

Qian, Y. Y., Zhang, H. N., Cui, S. F., Liu, Y., Wang, G. G., Jin, W. P., et al. (2017). Research progress of cotton pigment gland. *Crop Res.* 31, 205–210. doi: 10.16848/j.cnki.issn.1001-5280.2017.02.24

Qu, Y. F., Duan, Y. H., Kang, N. N., Du, H. Y., Cheng, S., and Huang, J. L. (2017). Genetic analysis of near-isogenic lines with different pigment glands on cotton. *J. Shanxi Agr Sci.* 45, 1–5. doi: 10.3969/j.issn.1002-2481.2017.01.01

Ronneberger, O., Fischer, P., and Brox, T. (2015) *U-Net: Convolutional networks for biomedical image segmentation. arXiv [Preprint]*. Available at: https://arxiv.org/abs/1505.04597v1 (Accessed May 18 2015).

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vision* 77 (1-3), 157–73. doi: 10.1007/s11263-007-0090-8

Sun, C., Ai, Y. B., Wang, S., and Zhang, W. D. (2021). Mask-guided SSD for small-object detection. *Appl. Intell.* 51, 3311–3322. doi: 10.1007/s10489-020-01949-0

Tetila, E. C., Machado, B. B., Menezes, G. K., Oliveira, A. D., Alvarez, M., Amorim, W. P., et al. (2019). Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks. *IEEE Geosci Remote S* 17, 903–907. doi: 10.1109/LGRS.2019.2932385

u, Y. S., Ni, C. F., Li, W. C., Lee, I. H., and Lin, H. P. (2020). Applying deep learning algorithms to enhance simulation of large-scale groundwater flow in IoTs. *Appl. Soft Comput.* 92, 106298. doi: 10.1016/j.asoc.2020.106298

Wang, Y. H., Cai, X. Y., Zhang, Z. M., Zhou, Z. L., Wang, K. B., and Liu, F. (2018). Comparison of the distribution of pigment glands in four cultivated cotton leaves. *China Cotton* 45, 16–19. doi: 10.11963/1000-632X.wyhlf.20181217

Wilson, F. D., and Smith, J. N. (1976). Some genetic relationships between gland density and gossypol content in gossypium hirsutum l. *Crop Sci.* 16, 830–832. doi: 10.2135/cropsci1976.0011183X001600060023x

Zhang, J., Meng, Y. Z., and Chen, Z. P. (2021). A small target detection method based on deep learning with considerate feature and effectively expanded sample size. *IEEE Access* 9, 96559–96572. doi: 10.1109/ACCESS.2021.3095405

Zhao, T. L., Xie, Q. W., Li, C., Li, C., Mei, L., Yu, J. Z., et al. (2020). Cotton roots are the major source of gossypol biosynthesis and accumulation. *BMC Plant Biol.* 20, 88. doi: 10.1186/s12870-020-2294-9

Zhou, P., Ni, B. B., Geng, C., Hu, J. G., and Xu, Y. (2018). "Scale-transferrable object detection," in *in Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, 528–537. doi: 10.1109/CVPR.2018.00062

Zhu, Z., Liang, D., Zhang, S. H., Huang, X. L., Li, B. L., and Hu, S. M. (2016). "Traffic-sign detection and classification in the wild," in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, 2110–2118. (Las Vegas). doi: 10.1109/CVPR.2016.232

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fpls.2022.1025663

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Karansher Singh Sandhu,
Bayer Crop Science (United States), United States
Gongfa Li,
Wuhan University of Science and Technology, China
Alexandr Muterko,
Institute of Cytology and Genetics (RAS), Russia
Yang Lu,
Heilongjiang Bayi Agricultural University, China
Lingfeng Duan, Huazhong Agricultural University, China

*CORRESPONDENCE
Ping Liu
liuping@sdau.edu.cn
Xiang Li
lixiang@sdau.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 23 August 2022
ACCEPTED 13 December 2022
PUBLISHED 11 January 2023

CITATION
Song X, Liu L, Wang C, Zhang W, Li Y, Zhu J, Liu P and Li X (2023) Real-time determination of flowering period for field wheat based on improved YOLOv5s model. *Front. Plant Sci.* 13:1025663. doi: 10.3389/fpls.2022.1025663

COPYRIGHT
© 2023 Song, Liu, Wang, Zhang, Li, Zhu, Liu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Real-time determination of flowering period for field wheat based on improved YOLOv5s model

Xubin Song[1†], Lipeng Liu[1†], Chunying Wang[1], Wanteng Zhang[1], Yang Li[1], Junke Zhu[2], Ping Liu[1*] and Xiang Li[3*]

[1]College of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian, China, [2]School of Agricultural and Food Engineering, Shandong University of Technology, Zibo, China, [3]State Key Laboratory of Crop Biology, College of Life Sciences, Shandong Agricultural University, Taian, China

The flowering period is one of the important indexes of wheat breeding. The early or late flowering affects the final yield and character stability of wheat. In order to solve the problem that it is difficult to accurately and quickly detect the flowering period of a large number of wheat breeding materials, a determination method of flowering period for field wheat based on the improved You Only Look Once (YOLO) v5s model was proposed. Firstly, a feature fusion (FF) method combing RGB images and corresponding comprehensive color features was proposed to highlight more texture features and reduce the distortion caused by light on the extracted feature images. Second, the YOLOv5s model was selected as a base version of the improved model and the convolutional block attention model (CBAM) was adopted into the feature fusion layer of YOLOV5s model. Florets and spikelets were given greater weight along the channel and spatial dimensions to further refine their effective feature information. At the same time, an integrated Transformer small-target detection head (TSDH) was added to solve the high miss rate of small targets in wheat population images. The accurate and rapid detection of florets and spikelets was realized, and the flowering period was determined according to the proportion of florets and spikelets. The experimental results showed that the average computing time of the proposed method was 11.5ms, and the average recognition accuracy of florets and spikelets was 88.9% and 96.8%, respectively. The average difference between the estimated flowering rate and the actual flowering rate was within 5%, and the determination accuracy of the flowering period reached 100%, which met the basic requirements of the flowering period determination of wheat population in the field.

KEYWORDS

field, wheat, machine vision, YOLOv5, flowering period determination

1 Introduction

Flowering period reflects the growth of various crops. The flowering period of wheat determines the growing period of wheat, which marks the transition from nutrient consumption to nutrient accumulation, and then influences the final yield and the stability of traits (Wang et al., 2020; Velumani et al., 2020; Zhu et al., 2021). Different varieties of wheat have different flowering periods and flowering traits, and the flowering situation is also changed of the same wheat varieties grown at different locations (Liu et al., 2015). The genotypic performances of different varieties under such conditions as diverse environments and planting practices can be determined by recording the flowering period of wheat. At present, the flowering period of a large number of wheat breeding materials and germplasms is mainly determined by manual estimation of the proportion of florets to spikelets, which is labor-intensive, time-consuming, and subjective (Zhang et al., 2020). Therefore, it is necessary to study an efficient and accurate detection technology to replace manual labor and provide critical information for the development of wheat germplasms.

The technology of plant phenotype monitoring has been called the fourth revolution in agricultural production (Sishodia et al., 2020). Efficient, automated, and versatile phenotypic techniques are important tools for accelerating the breeding process and improving genetic gain (Zhao et al., 2019). Phenotypic acquisition built on machine vision is a fast, low-cost, non-destructive technique that can achieve accurate extraction of target traits (Hu et al., 2019; Ajlouni et al., 2020; Zhang et al., 2020). With the development of imaging equipment and image processing algorithms, machine vision technology has made tremendous advances in detecting plant flowering periods. At present, many research results on flowering detection of fruit trees, shrubs, and other plants. Xiong et al. (2021) realized the instance segmentation of lychee flowers based on the DeepLab V3 network model. The recognition accuracy of the network model reached 87.0%, and the detection time of a single image was 67ms. Zhao et al. (2020) proposed a tomato flowering detection method based on the cascading convolutional neural network to identify tomato flowers at different flowering stages. The average recognition accuracy of the proposed method was 82.8%, and the detection time of a single image was 12.5ms. Most of the above studies were conducted on fruit trees and shrub plants with low planting density and large differences in flower characters, leaves and background. In a complex field environment, Milicevic et al. (2020) solve the problem of mutual occlusion among corn plants by installing hundreds of automatic cameras in the experimental field to collect images parallel to the top of each corn. In order to realize the detection of corn flowering, the main spike of the tassel was segmented based on the image processing method, and the recognition accuracy of complex targets was improved by using the cross-entropy loss function of dynamic scaling. The average recognition accuracy of the maize tassel was improved to 91.1%. Cai et al. (2021) proposed a method to determine the flowering period of the sorghum based on multi-temporal spike counting to realize flowering period detection in the field environment. The YOLOv5 model was used to detect and count sorghum spikes, and the average recognition accuracy was 86.2%. The proposed method was able to accurately detect sorghum spikes and calculate the flowering period. However, the above studies did not rule out the possibility that the influence of light conditions on the determination. When the reflection of the wheat spike was obvious, the recognition accuracy would be affected.

Over the years, scholars have explored some computer vision detection methods to study the flowering period of wheat. Sadeghi-Tehran et al. (2017) presented an automated method to detect wheat heading and flowering stages. The bag-of-visual-word technology was used to identify the flowering of wheat ears in digital images and determine whether the wheat in the image is in the flowering stage. The accuracy rate of flowering detection was 85.45%. Ma et al. (2020) proposed a two-stage segmentation method based on superpixel clustering and the fully convolutional network (FCN) to realize the segmentation of wheat spikes of the wheat canopy image at the flowering period. The accuracy of flowering spikes segmentation was 83.7%. The above studies were to realize the recognition of flowering wheat by importing color and texture features into the Support Vector Machine (SVM) for training, or using the convolutional neural network to train labeled flowering ears and non-flowering ears. However, the existing research only identified the flowering and non-flowering wheat ears and did not determine the flowering period of wheat in the field, which could not provide objective and comprehensive data.

The flowering period of wheat in the field is mainly determined by the proportion of florets to spikelets. Compared to other crops, the florets and spikelets of wheat have small morphological structures and the color differences are not obvious (Bommert et al., 2005; Cheng et al., 2016), which increased the difficulty of detection. In particular, there are the following difficulties to realize flowering time detection of wheat populations in the field environment: Firstly, the environmental background is complicated, and common image pretreatment methods cannot globally suppress noise from light, wheat awns, leaves and soil; secondly, when wheat enters the flowering period, the leaf extension is large, and the spikelets will overlap with each other. Finally, florets and spikelets belong to small-scale targets in the population of images, and the scale changes drastically according to the different shooting distances.

Therefore, to achieve an accurate determination of the flowering period in wheat, firstly, we proposed an image enhancement method based on feature fusion, which reduces the light distortion of wheat images and highlights more texture features. On this basis, an improved YOLOv5s model was proposed to optimize the extraction efficacy of obscured wheat

spikes, and solve the problem of the high detection rate of small target leakage in wheat population images. Finally, the accurate and fast detection of florets and spikelets was achieved, and the ratio of florets to spikelets was used to determine the flowering period. It can provide data support for wheat stable yield improvement.

2 Materials and methods

2.1 Data acquisition

The images were taken at the Agricultural Experimental Station in Shandong Agricultural University, Taian, China [36° 9′52″N, 117°9′21″E]. During the shooting period, the weather was mostly sunny, with less cloudy and rainy days. The image data were acquired by the data acquisition platform as shown in Figure 1A. The Jierui Weitong DW800_2.9mm camera(Lens: no distortion, wide angle 2.8mm; Angle: Oblique 45 degrees; Shutter speed: 1s) was mounted on the side of the Phenotype platform, 1m above the ground. The camera manufacturer is Shenzhen Jerui Weitong Electronic Technology Co., LTD, and the origin is Shenzhen, China. Subsequently, the acquired image of 4000x3000 pixels was clipped to 800x600 pixels to improve the operation efficiency, and the middle five images were retained to avoid blurring of florets and spikelets. The cropping method is shown in Figure 1B. To ensure the accuracy of the experiment, 4570 wheat images of wheat including different varieties (including TKM33, SN48, JM44, and SN27), flowering periods, planting densities, shooting angles, weather conditions, and light intensities were collected.

2.2 YOLOv5 model

The convolution neural network (CNN), as a neural network based on the principle of biological visual neural perception, includes convolution computations and a kind of feedforward neural network with a deep structure (Chen et al., 2021; Li et al., 2021). Object detection methods based on deep learning are developing rapidly and can be roughly divided into two types: the two-stage detection method based on candidate region and the one-stage detection method based on regression (Huang et al., 2021; Yun et al., 2022). Among them, the YOLO (You Only Look Once) series was an essential part of one-stage detectors (Redmon et al., 2016). In the task of target detection, YOLO could obtain global-context information by looking at the input images only once.

YOLOv5 is one of the YOLO series models, which achieves a better balance between accuracy and speed than the previous version. The YOLOv5 framework is composed of input, feature extraction module (Backbone), feature fusion module (Neck), and output (Prediction). The backbone network is responsible for feature extraction targets, and the Neck network generates feature pyramids for object scaling. The prediction network adopts three scales of head: small (80×80×128), medium (40×40×256), and large (20×20×512) for final detection. YOLOv5 is divided into YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x models according to the depth and width of the network. The overall structure of these five models is the same. The width of the network determines the number of convolution kernels, which is the learning ability of the network to extract features. The depth of the network determines the number of components at each level, that is, the ability to fuse

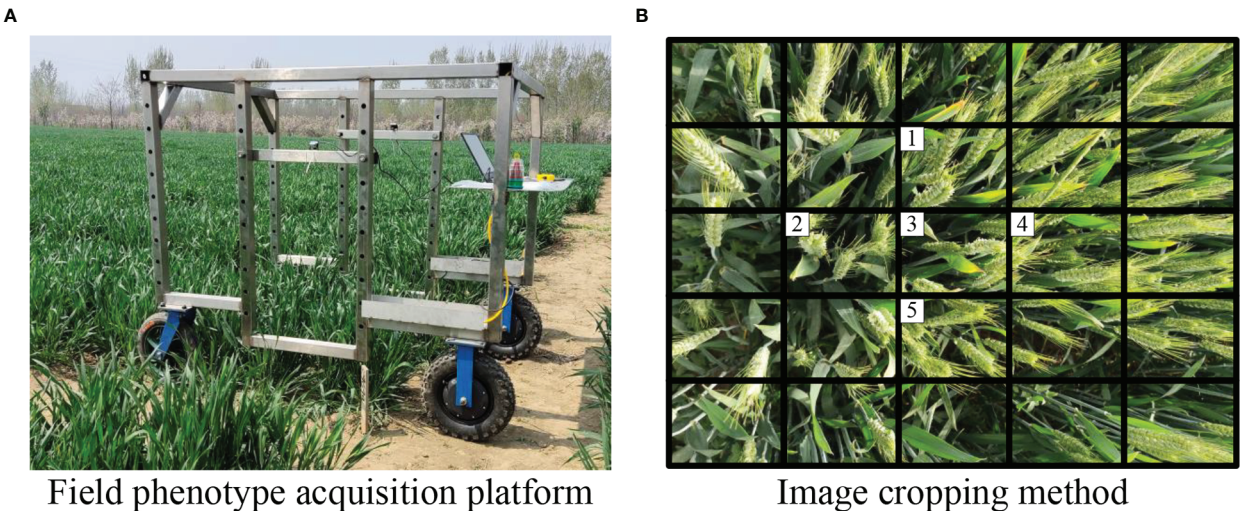


FIGURE 1
Wheat image data acquisition. **(A)** Field phenotype acquisition platform. **(B)** Image cropping method.

features and the speed of model convergence. Considering that this paper is applied to the detection of the wheat flowering period in the field with high real-time requirements, a network model based on improved YOLOv5s is proposed.

2.3 Determination method of flowering period based on improved YOLOv5s

The determination method of the wheat flowering period is mainly divided into two steps. The first step is to accurately identify florets and spikelets based on the improved YOLOv5s model. Then the ratio of florets to spikelets was used as the flowering rate. When the flowering rate was over 50%, the time of image capture was read and recorded to complete the determination of the flowering period. The determination process of the flowering period is shown in Figure 2.

There are three major improvements from the original YOLOv5s to the improved model are described below: (1) A feature fusion (FF) method that combining RGB images with corresponding comprehensive color features (CCF) was proposed, attenuating the distortion of the extracted feature image caused by light and highlighting more texture features. (2) The convolutional block attention module (CBAM) (Woo et al., 2018) was inserted into the original YOLOv5s network. The CBAM assigned a large weight to floret and spikelet features by combining the channel attention module and spatial attention module. It can learn target features well and suppress non-targets features to improve detection accuracy. (3) Considering that the target scale changed dramatically and a large number of small targets in the wheat dataset, an integrated Transformer small-target detection head (TSDH) was added to combine with the other three detection headers of the original YOLOv5s network for accurate spikelet and floret detection.

2.3.1 Feature fusion method

Compared with rice, corn, and other crops, the background of wheat population images in the field is more complex, the morphological structure of florets and spikelets is small, and the color difference between them is also not palpable (Bommert et al., 2005; Feng et al., 2022; Huang et al., 2022). Converting RGB images to common color spaces such as HSV or Lab does not solve the global noise caused by light, wheat awns, leaves, and soil. Therefore, in the previous study, a comprehensive color feature (CCF) method that can be adjusted adaptively according to the light intensity and clarity of images was proposed by comparing the data feature of different color spaces and different color indices (Liu et al., 2022). In this study, the adaptive adjustment CCF method was used to reduce the influence of light and enhance the differences of florets, spikelets and other targets.

As shown in Figure 3A, the accuracies of the original YOLOv5s model in identifying florets and spikelets were only 60.7% and 80.3%. The detection results of the original YOLOv5s model showed that the main reason for the low recognition accuracy is a relatively uniform degree of feature standard extracted from RGB images. However, different wheat varieties have different traits, and there is a lack of high-weight added-value eigenvectors that can highlight the characteristics of florets and spikelets in the process of model classification. Due to the small floret target and high noise in the field environment, the original YOLOv5s model would greatly increase the probability of missing recognition and misidentification when raising the detection threshold. This made a huge difference between precision and recall. The characteristics of florets and leaf reflective spots are similar under natural conditions, which led to misrecognition and the reduced detection ability of the original YOLOv5s model for florets. As shown in Figure 3B, wheat leaves were mistaken for florets. Therefore, more

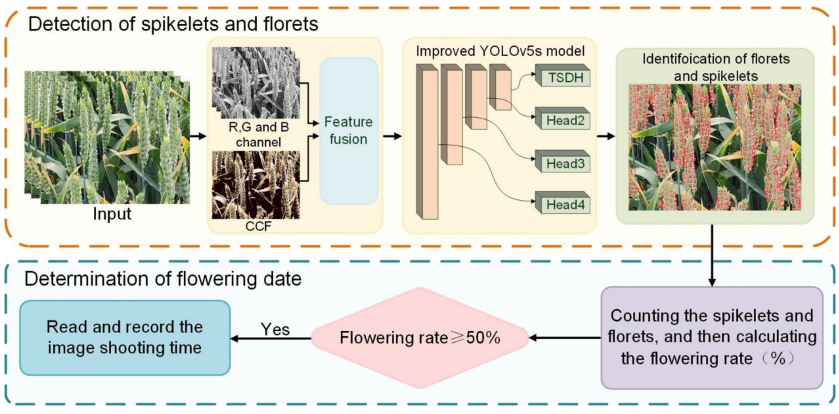


FIGURE 2
Determination process of wheat flowering period.

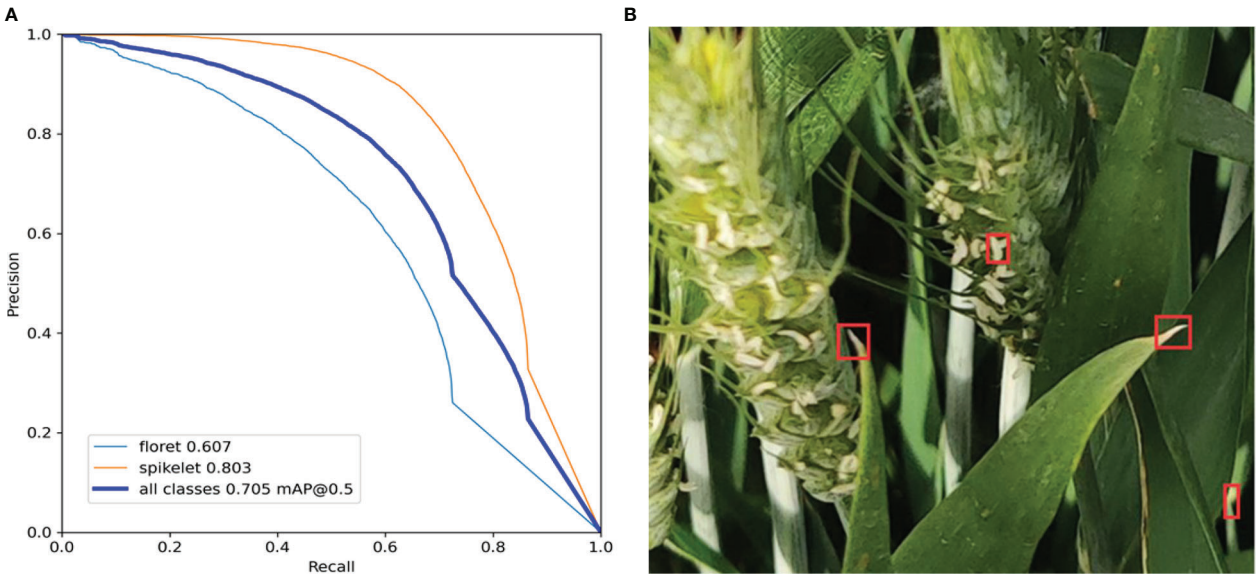


FIGURE 3
Identify situation of the Original YOLOv5s model. **(A)** precision-recall curve of florets and spikelets(floret 0.607 mAP@0.5, spikelet 0.803 mAP@0.5). **(B)** misidentification of florets(red box).

dominant traits need to be provided decision support for the detection layer.

Figure 4 shows that both shallow and deep features after convolutional extraction have distortion features (red marked box) caused by the light and redundant features (green marked box) that increase model convergence time during training. To solve the above problems, the main methods available are to reduce the number of convolution cores or to fuse images in multiple sources to increase the extraction of effective features. However, the former is suitable for situations where the background of image data is highly controllable, such as indoor environment or image data collected through a standardization process, but not for complex field environments (Liu et al., 2021; Zhao et al., 2021; Bai et al., 2022).

Therefore, a feature fusion (FF) method combing RGB images and corresponding CCF was presented to solve the above difficulties, and subsequent feature extraction and feature fusion are performed based on the fused images. The feature fusion method workflow, illustrated schematically in Figure 5, mainly contains three steps: (1) The R, G, and B

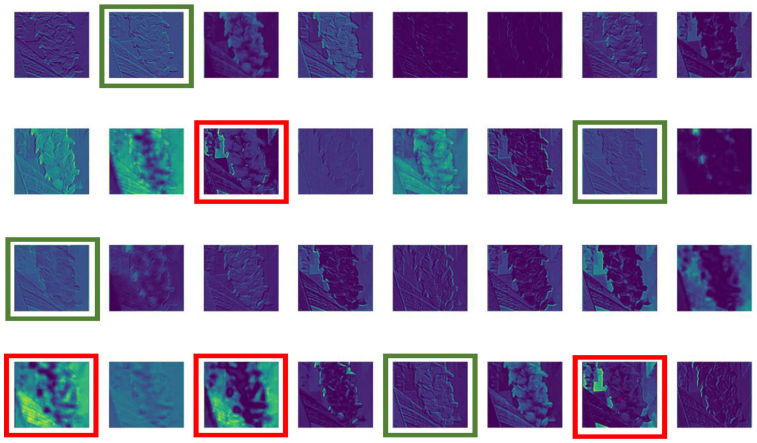


FIGURE 4
Partial feature map extracted by the Original YOLOv5s model (Red marked box: distortion features. Green marked box: redundant features).

channels in RGB images and the channel information of the CCF are extracted and transmitted to the input layer of the convolutional neural network. (2) In the feature extraction stage of the YOLOv5 model, the shallow and deep features of the two image sources input in step 1 are extracted. (3) The input features are addressed by independent convolution, pooling, and full connection networks, and the extracted features are stitched with equal weight.

In Figure 5, the main function of the convolution layer is to extract local features from the pixel information of input images. The scale of features extracted by the convolution kernel is relatively large, and using the above features for target classification will produce a large amount of computation and affect the inference speed. Therefore, secondary extraction of image features using a pooling layer reduces the feature parameters. Considering that the background of the field image is complex and noisy, pooling the extracted features with kernel sizes of 3, 5, and 7 to better preserve the feature texture and improve the generalization of the model. The same gradient descent algorithm as pre-training was used for model training, and the parameters of the convolution layer and pooling layer in the convolutional neural network were updated by back propagation.

Compared with Figure 4, the proposed feature fusion method attenuated the distortion of the extracted feature image caused by light and highlights more texture features. It is likewise easier to classify florets and spikelets in shallow features with clearer details, and the results are shown in Figure 6.

2.3.2 Convolutional block attention module

The complementarity between RGB images and corresponding CCF was enhanced by the feature fusion method in the previous section. The feature fusion method eliminated the influence of background noise and extracted more characteristics of florets and spikelets in complex environments, but increased the input image from three to four dimension. The attention mechanism was integrated into the YOLOv5s model to improve the model, could make the network more focused on spikelets and florets, and avoid too many features affecting the computational power and convergence speed of the model. The Convolutional block attention module (CBAM) is one of the most effective attention mechanisms and consists of channel attention module and spatial attention module. It can enhance the ability of the model to extract image features and suppress invalid background information by redistributing the originally uniform distribution resources according to the importance of the detection target. So, the CBAM was adopted into the feature fusion layer of YOLOV5s model. The calculation process of the CBAM is shown in Equation (1).

$$\begin{aligned} F' &= M_C(F) \otimes F \\ F'' &= M_S(F') \otimes F' \end{aligned} \tag{1}$$

In the above formula, F represents the input eigenvector, MC is the channel attention feature map, F' represents the channel attention module outputs feature vectors, MS is the spatial attention feature map, and F'' represents the spatial attention module outputs feature vectors.

CBAM links the channel attention module and spatial attention module in series. The channel attention module first compressed the input feature map by average pooling and maximum pooling, and then sent it to the shared multi-layer perceptron (MLP) structure for processing. Finally, the channel attention map was generated by the sigmoid function to solve the problem of what is the target. The spatial attention module used convolution and the sigmoid function to process the input feature map and ultimately determined where to pay attention.

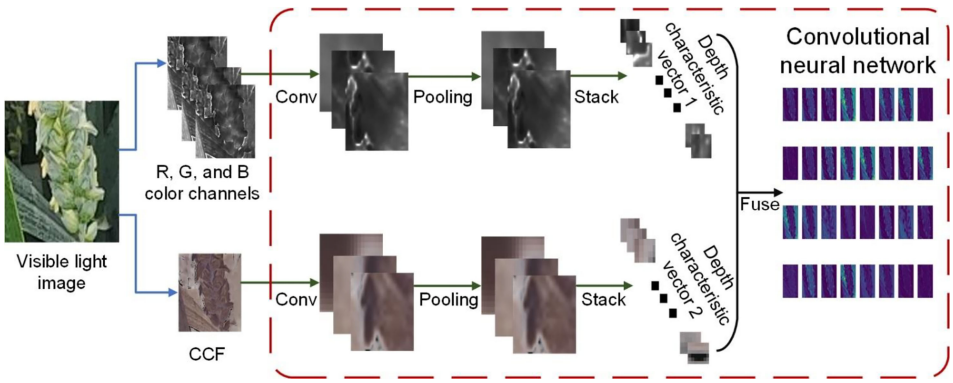


FIGURE 5
The overall schematic diagram of feature fusion method.

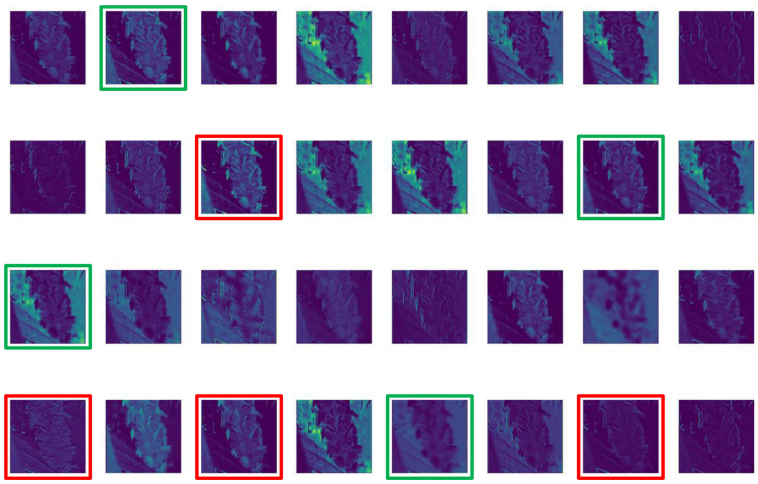


FIGURE 6
Partial feature map extracted based on feature fusion method.

Compared with separate channel attention network SENet and spatial attention network STN, the CBAM does not increase too much computation (Zhao et al., 2022). It is a lightweight module that can be integrated into the most well-known CNN architecture and can be trained in an end-to-end manner. The CBAM only needs to give a feature mapping at the convolution layer, then it will infer attention mapping in turn along two independent dimensions, channel, and space. The attention map is then multiplied by the input feature to perform adaptive feature refinement. Therefore, the CBAM is a simple but efficient attention module, and its module structure is shown in Figure 7.

The CBAM was added after the C3 module of the neck to update feature mapping weights after each residual convolution in the feature fusion stage. It can effectively improve the accuracy of the improved YOLOv5s in florets and spikelets target detection by weighting the feature map weights of different channels and spatial dimensions.

2.3.3 Multi-detection heads structure integrated transformer

After normalizing the coordinate dimensions in the acquired dataset, the length and width size distributions and pixel area distribution of each labeled target were collected for statistics. The result is shown in Figure 8. Without data enhancement, the pixel area of 9325 labeled boxes is less than the minimum detection box size of YOLOv5s, accounting for about 10% of all labeled boxes.

It was found that the whole image of the wheat field contained many tiny examples due to the small size of florets and spikelets analyzing the labeled data. So, an up-sampling operation was inserted into the neck of YOLOv5s and a detection head was embedded to detect tiny targets. The improved microscale layer generated a feature map by extracting the underlying spatial features and fusing them with deep semantic features, which made the network structure more

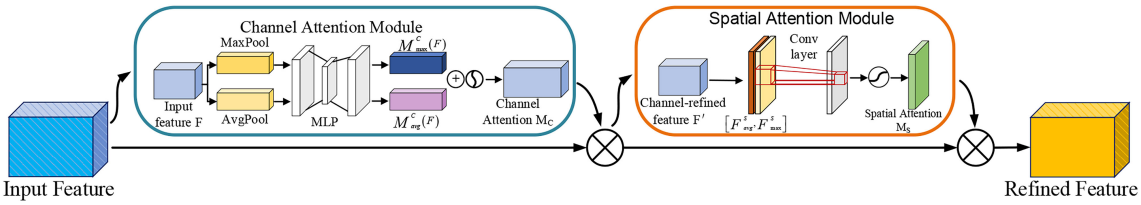


FIGURE 7
The overall structure of convolutional block attention module: MaxPool represents the maximum global pooling; AvgPool represents global average pooling; MLP represents a multi-layer perceptron with shared weights; Conv indicates convolution operation.

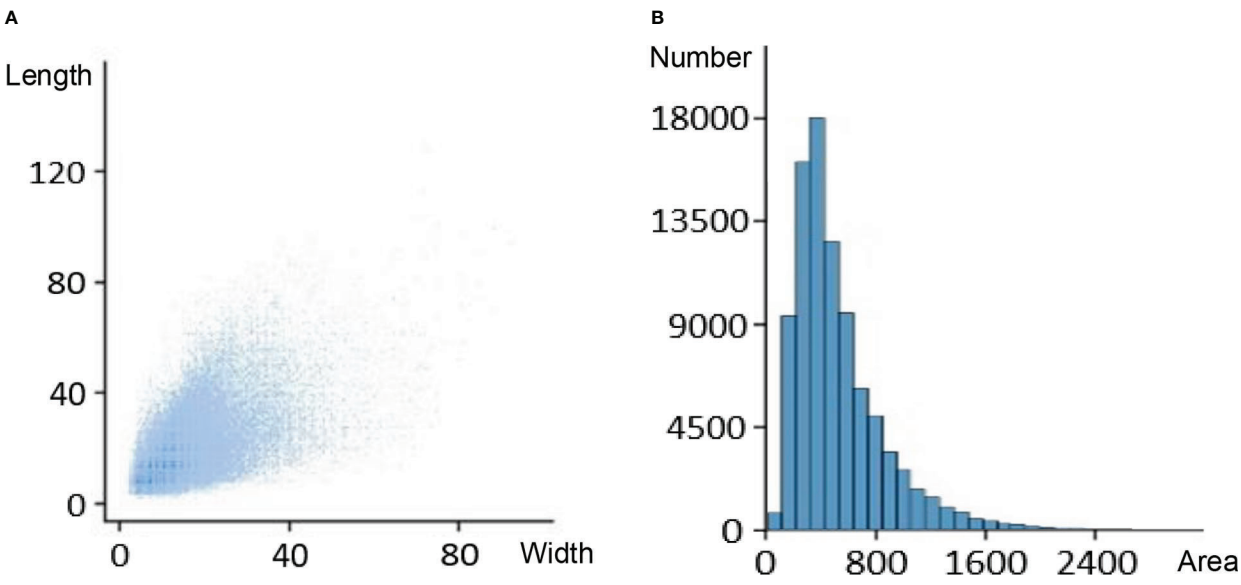


FIGURE 8
Labeled data distribution: **(A)** size distribution. **(B)** area distribution.

extensive and detailed. Combined with the other three detection heads, the improved YOLOv5s four-head structure can alleviate the negative effects caused by the large size distortion of florets and spikelets.

The newly added detection head is located at the end of the neck of YOLOv5s. After multiple up-sampling, the resolution of the feature map is lower, and blurred small targets will be missed. Inspired by Vision Transformer (Dosovitskiy et al., 2020), the C3 module originally connected to the newly added detection head was replaced with the Transformer encoder module shown in Figure 9. In the first step, the feature map with the input scale of W×H was divided into 1 × 1 pixel blocks, and the positions of each pixel block in the feature image were sequentially entered into the encoder. Transformer encoders are composed of two sub-layers. The first is the multi-head attention layer, and the second (MLP) is the fully-connected layer. Each sub-layer is connected using residual connections. To improve model generalization and reduce computation costs, the dropout layer is behind each sublayer. Different eigenvectors in the encoder containing spatial location information were transferred to the dropout layer through the attention module of the feed-forward neural network. The dropout layer discarded the low-weight features and then transmitted the valid features to the full connection layer to complete the classification task.

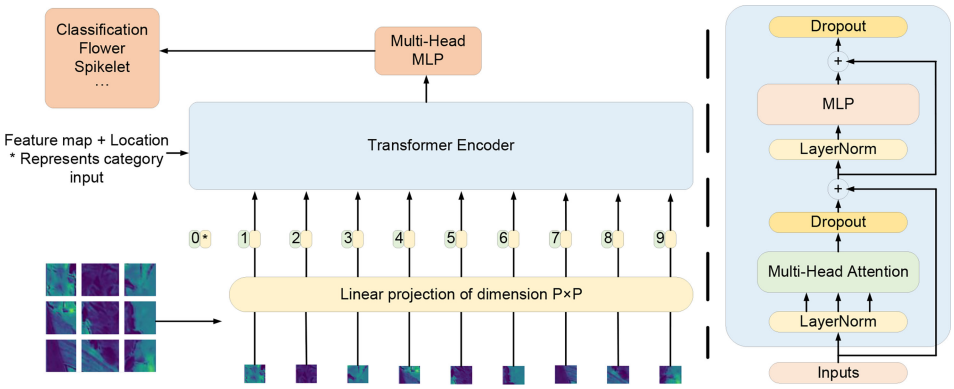


FIGURE 9
The diagram of transformer encoder structure.

The ability to capture global information and upper-level features can be enhanced by integrating Transformer small-target detection header. Furthermore, it can exploit the self-attention mechanism to explore the potential of feature representation (Zhu et al., 2021). The overall structure of the improved network is shown in Figure 10.

Multiple convolution and up-sampling operations are required while adding a detection head. To avoid slow model convergence and gradient explosion caused by the increase of feature dimensions, the maximum pooling layer (SPP) was embedded in front of the second, third, and fourth detection heads. The added maximum pooling layer can ensure the integrity of the overall image features while reducing parameters. Table 1 presents the parameters of the improved YOLOv5s model.

2.3.4 Determination of flowering period

The actual flowering period could not be accurately determined by the ratio of the flowering spikes to the no-flowering spikes. Because one flower or ten flowers on a single wheat spike can be counted as a flowering spike, the true flowering condition of the single wheat spike cannot be fully judged, thus affecting the determination of the wheat flowering period. The agronomic criterion for the current wheat flowering period identification is that the inner and outer glumes of the florets in half of the spikelets in the plot are opened and the pollen is dispersed. At present, the flowering period of a large number of wheat breeding materials is mainly determined by manual estimation of the proportion of florets to spikelets (Zhang et al., 2020; Wang et al., 2020). So this study determined the flowering period of a plot based on the flowering rate, which is the ratio of the spikelets to florets in all images obtained by the community. When the flowering rate exceeded 50%, the plot was determined to be in flowering period, otherwise, it was determined to be heading period.

2.5 Experimental training and evaluation indicators

The dataset in this paper is composed of spikelets and florets, and the field environment is relatively complex. The effect will be better if there are public datasets with similar detection tasks to use the transfer learning training model. However, no similar detection task dataset was found at present, so transfer learning was not performed. For a fair comparison between the models, each model was trained from scratch. The learning rate was based on cosine annealing attenuation strategy (Loshchilov and Hutter, 2016), as shown in Equation (2), and the number of iterations was 300.

$$n_t = n_{\min}^i + \frac{1}{2}\left(n_{\max}^i - n_{\min}^i\right)\left(1 + \cos\left(\frac{T_i}{T}\Pi\right)\right) \tag{2}$$

In the above formula, $n_{\min}^i$, $n_{\max}^i$ represent the minimum and maximum of the i-round learning rate, T_i represents the cumulative number of samples during the i-round training, and T represents the total number of samples.

The comprehensive recognition accuracy and real-time performance of the model under the five categories of florets, spikelets, and background were measured using three indicators: precision P, recall rate R, *F1-score*, *accuracy*, and mean average precision *mAP*, as shown in Equations (3-7).

$$P = \frac{TP}{TP + FP} \tag{3}$$

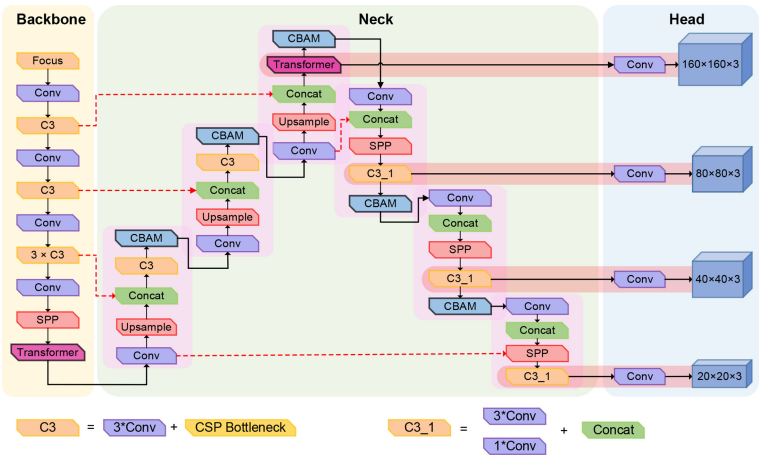


FIGURE 10
Schematic diagram of improved YOLOv5s network structure.

TABLE 1 Improved layer parameters of YOLOv5s network.

Layer	Module	stack	Output size	Source	Input dimension	Output dimension	Convolution
0	Focus	1	320	-1	3	32	3×3
1	Conv	1	160	-1	32	64	3×3
2	C3	1	160	-1	64	64	[1, 3]
3	Conv	1	80	-1	64	128	3×3
4	C3	1	80	-1	128	128	[1, 3]
5	Conv	1	40	-1	128	256	3×3
6	C3	3	40	-1	256	256	[1, 3]
7	Conv	1	20	-1	256	512	3×3
8	SPP	1	20	-1	512	512	[3, 5, 7]
9	Trans	1	20	-1	512	512	—
10	Conv	1	20	-1	512	256	1×1
11	Upsample	1	40	-1	256	256	—
12	Concat	1	40	-1, 6	256	512	—
13	C3	1	40	-1	512	256	[1, 3]
14	CBAM	1	40	-1	256	256	—
15	Conv	1	40	-1	256	128	1×1
16	Upsample	1	80	-1	128	128	—
17	Concat	1	80	-1, 4	128	256	—
18	C3	1	80	-1	256	128	[1, 3]
19	CBAM	1	80	-1	128	128	—
20	Conv	1	80	-1	128	64	1×1
21	Upsample	1	160	-1	64	64	—
22	Concat	1	160	-1, 2	64	128	—
23	Trans	1	160	-1	128	128	—
24	CBAM	1	160	-1	128	128	—
25	Conv	1	80	-1	128	64	3×3
26	Concat	1	80	-1, 20	64	128	—
27	SPP	1	80	-1	128	128	[3, 5, 7]
28	C3_1	1	80	-1	128	256	[1, 3]
29	CBAM	1	80	-1	256	256	—
30	Conv	1	40	-1	256	128	3×3
31	Concat	1	40	-1, 15	128	256	—
32	SPP	1	40	-1	256	256	[3, 5, 7]
33	C3_1	1	40	-1	256	512	[1, 3]
34	CBAM	1	40	-1	512	512	—
35	Conv	1	20	-1	512	256	3×3
36	Concat	1	20	-1, 10	256	512	—

(Continued)

TABLE 1 Continued

Layer	Module	stack	Output size	Source	Input dimension	Output dimension	Convolution
37	SPP	1	20	-1	512	512	[3, 5, 7]
38	C3_1	1	20	-1	512	1024	[1, 3]
39	Detect	1	20,40,60,120	23,28,33,38	—	—	—

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$F1 - score = \frac{2PR}{P + R} \tag{5}$$

$$mAP = \frac{1}{c}\sum_{k=i}^{N} P_{(k)} R_{(k)} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

Among them, *TP* represents the number of spikelets and florets correctly identified by the model, *FP* represents the number of false recognition of the background as spikelets and florets, *FN* represents the number of spikelets and florets not identified, *TN* represents the number of background correctly identified by the model, *C* represents the sample category, *N* represents the threshold of citations.

The higher the value of *P*, *R*, and *mAP*, the higher the accuracy of object detection, and the average run time (ms/frame) was the average time the model takes to process the wheat image.

3 Results and discussion

The florets identification and the determination of the flowering period of wheat populations in the field have yet to be studied. To verify the effectiveness and adaptability of the improved YOLOv5s model under different conditions, quantitative and qualitative tests were conducted on the improved YOLOv5s model and the original YOLOv5s model in the test set images. And the performance differences between the improved model and other models were studied by comparing with other advanced and non-deep learning methods. Afterward, an ablation experiment was conducted to explore the optimization effect of various improvement strategies applied in the YOLOv5s model, enabling better detection capability in this work. Finally, the actual performance of the model for flowering determination of wheat population in the field was further tested, and the field phenotype acquisition platform was invoked as the carrier to conduct field experiments.

3.1 Quantitative test

From 457 images in the test set, 20 images with different panicles, different varieties, shooting angles, and different planting densities were selected for calculation and analysis as shown in Figure 11. Florets and spikelets were counted by manual method and model respectively, and the classification accuracy was calculated. The statistical results are shown in Table 2. The confidence threshold was placed at 50%. The original YOLOv5s model is represented by model 1, and the improved YOLOv5s model is represented by model 2.

Table 2 shows that the improved YOLOv5s model and the original YOLOv5s model have high recognition accuracy in images with low planting density and low background noise (such as No. 2, 5, 12, and 16). Comparing images No. 4, 7, and 8, when the degree of occlusion is small and the target is obvious, the original YOLOv5s model has poor adaptability to the distortion of wheat morphology. The improved YOLOv5s model has higher detection accuracy for distorted florets and spikelets, and the recognition accuracy rate was 87.7% and 91.5%. When the influence of light is strong (such as images No. 6, 14, and 19), the recognition accuracy of the original model for florets and spikelets was only 63.9% and 81.2%, and the recognition accuracy of the improved model was 83.7% and 94.1%. For wheat images with large differences in characters (such as images No. 1, 11, 17, 18, and 20), compared with the original YOLOv5s model, the recognition accuracy of florets and spikelets by the improved YOLOv5s model was improved by 12.8% and 8.3% respectively.

For a more intuitive reflect the recognition effect of the model, the evaluation criteria for 457 images in the test set are shown in Table 3. The improved YOLOv5s model has a detection precision of 95.3%, recall rate of 86.2%, *mAP* of 92.9%, and an average detection time of 11.5ms for a single image. Although the detection time was increased by 38.8% compared with the original YOLOv5s model, the number of video frames detected per second was 86Fps, which is suitable for the current mainstream visible light cameras.

3.2 Qualitative tests

The effectiveness and adaptability of the improved YOLOv5s model under various conditions were proved by quantitative

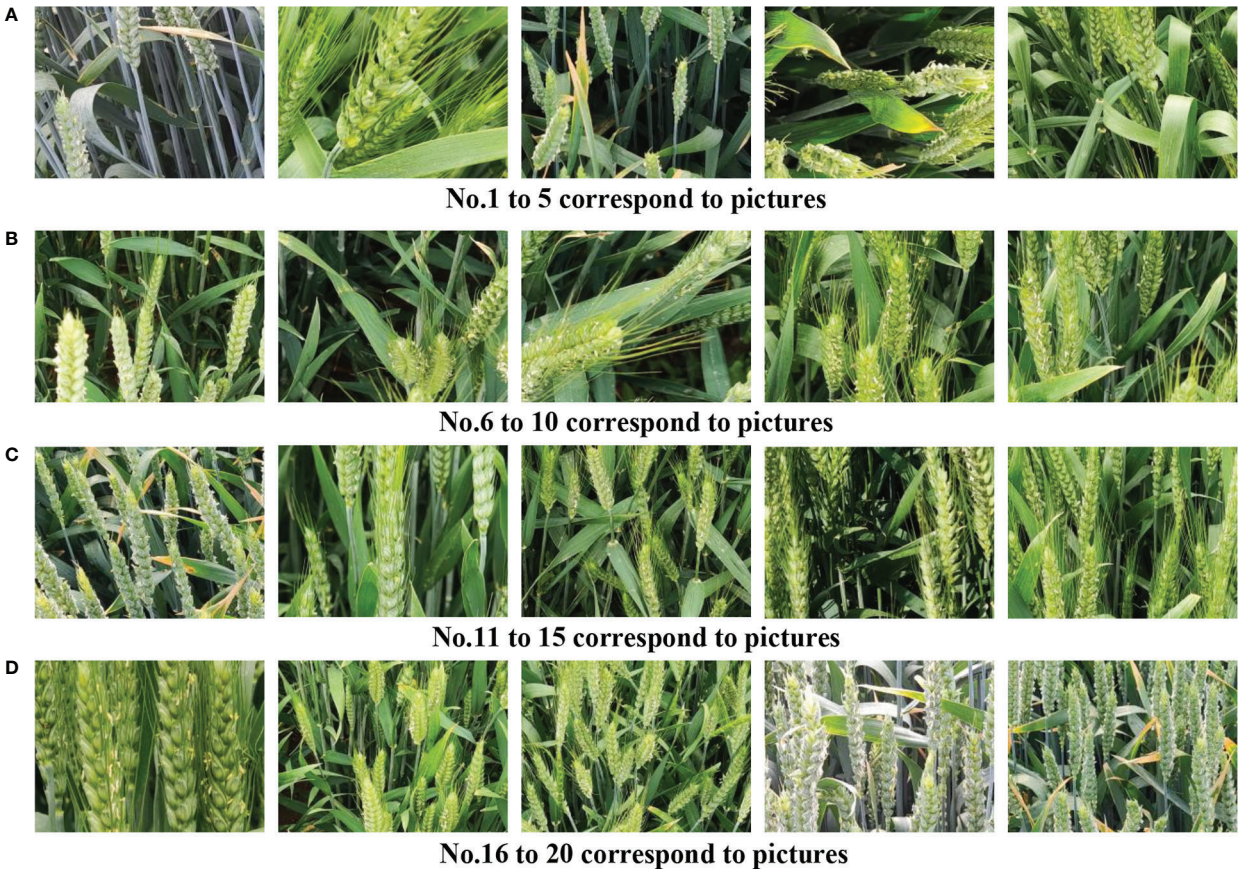


FIGURE 11
Twenty images with different numbers of spikes, different varieties, different shooting angles and planting densities. **(A)** The numbers 1 to 5 correspond to the picture. **(B)** The numbers 6 to 10 correspond to the picture. **(C)** The numbers 11 to 15 correspond to the picture. **(D)** The numbers 16 to 20 correspond to the picture.

experiments. In order to test the recognition effect of the model in various situations more intuitively, some images were selected for qualitative tests. As shown in Figure 12, this section divides some photos in the test set into five categories for qualitative tests: strong influence of light effect, severe angular distortion, blurred target area, serious occlusion, and population phenotype enrichment. All of the above are important factors affecting the robustness of the model in field image detection (Hu et al., 2019; Ajlouni et al., 2020).

Under the influence of strong light, the surface reflection of spikelets is serious, and the visual effect at the texture and endpoint of the leaves is similar to the color characteristics of the floret, which can easily cause misrecognition. The recognition situation when the local or global light is strongly affected is shown in Figure 12A. Only once had the leaf been misidentified as the spikelet, as indicated by the blue box in the figure. It can be seen from the recognition effect that the improved model had good adaptability to light, and there was no missing detection of wheat spikes and false detection of florets due to light reflection.

Wheat at the flowering period just shifts from vegetative growth to reproductive growth, the stem is soft and the accumulation of dry matter in the spike increases, causing the wheat to show signs of lodging. The angle or shape distortion of the collected images caused by the skew of the spikelet is shown in Figure 12B. According to the results, the improved YOLOv5s model had a beneficial effect on the detection of floret targets under distortion. However, when the tilt angle of spikelets is large, the spikelets are too dense, leading to a small number of spikelet targets with large distortion would be missed. The black box in the figure shows missing spikelets.

Phenotypic acquisition platform with large bumps and undulations during driving, as well as excessive wind speed during photographing, would cause blurring of the collected image information. Shown in Figure 12C is the recognition effect of fuzzy targets, in which the model prediction performance was

TABLE 2 Counting statistics results.

No.	Number of florets			Number of spikelets		
	Manual count	Model 1	Model 2	Manual count	Model 1	Model 2
1	48	34	39	57	50	54
2	7	5	5	46	40	43
3	69	38	47	121	94	106
4	52	31	43	107	86	99
5	15	12	13	94	88	98
6	43	28	36	129	117	128
7	38	24	35	50	39	45
8	32	22	29	44	34	40
9	80	42	73	86	73	79
10	52	38	44	110	95	107
11	139	87	103	226	192	211
12	16	11	11	62	55	59
13	36	21	27	134	118	132
14	38	22	33	115	98	112
15	81	54	72	148	117	137
16	52	37	48	68	66	66
17	58	31	42	216	173	190
18	91	67	78	273	237	256
19	146	95	121	214	159	191
20	188	131	155	335	276	309

(Model 1 represents the original YOLOv5s model. Model2 represents the improved YOLOv5s model.).

better for distant blurred targets but less effective for near targets with larger target sizes. By detecting the fuzzy images in the test set, the recognition accuracy of florets and spikelets was 86.4% and 90.7%, respectively, reduced by 2.5% and 6.1% compared with the average recognition accuracy of the improved model. The improved YOLOv5s model had no significant reduction in recognition accuracy due to image blurring and has good adaptability to target blurring.

Different wheat varieties have distinct traits. It was noted that some varieties of wheat had long and dense awns, which was the main factor to produce image noise by comparing different wheat varieties. Some wheat varieties had a serious overlap of spikes due to the high number of tillers. The recognition of serious occlusion overlaps is shown in Figure 12D. As shown in Figure 12D-2, the improved YOLOv5s model had higher accuracy in identifying the wheat varieties with severe overlap. However, in Figure 12D-1, the noise produced by over-dense wheat awns of this variety caused the hidden florets and spikelets to be missed detection. The black box in the figure shows the missed detection of florets and spikelets.

TABLE 3 Detection results of improved YOLOv5s model test on the test dataset.

Models	Recognition accuracy	Running time/s
YOLOv5s	$P = 0.719$ $R = 0.637$ $mAP = 0.705$	3.79
Improved YOLOv5s	$P = 0.953$ $R = 0.862$ $mAP = 0.929$	5.26

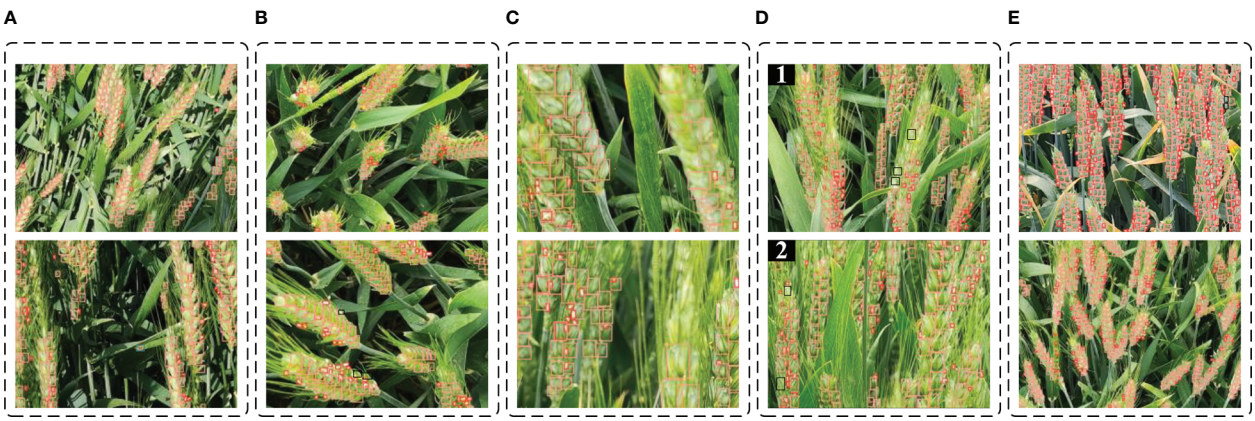


FIGURE 12
Detection effect of improved YOLOv5s model in different complex situations (The pink and dark red boxes represent true positive; black boxes represent false negative; blue boxes represent false positive). **(A)** strong light. **(B)** angle distortion. **(C)** blurred target area. **(D)** severe occlusion. **(E)** population phenotype enrichment.

In the plots with high planting density, the acquired images contain rich phenotypic information of the population, as shown in Figure 12E. The area of pixels occupied by the region of interest would tend to be less and is prone to small target missing when the image contains too much phenotypic information. The improved YOLOv5s model can accurately determine the region of interest and detect florets as well as spikelets, as seen by the identification effect.

3.3 Comparative experiments of different object detection algorithms

This study was compared with other advanced deep learning and non-deep learning methods to investigate the performance differences between the improved YOLOv5s model and other models. To guarantee the reliability of the results, the six networks(i.e., YOLOv3, YOLOv7, Faster R-CNN, Cascade R-CNN, superpixel segmentation, Improved YOLOv5s) were trained using the training and validation datasets in the same training environment. The training results are shown in Table 4, and the improved YOLOv5s achieved the best performance in all indicators, including the accuracy of florets and spikelets was 88.9% and 96.8%, F1-score of 90.52%, and mean average precision of 92.9%. Taken together, the improved YOLOv5s method presented had the best performance for detecting florets and spikelets compared with other detection methods, which proved the validity of the model proposed in this paper.

3.4 Ablation studies

In the previous section, quantitative and qualitative tests were carried out for the detection of the improved YOLOv5s model under various complex conditions and a quantitative comparison was made with the detection result of the original YOLOv5s model in the test set. The validity and adaptability of the improved model were proved. As mentioned earlier, three major improvements were made to the original YOLOv5s, including image enhancement (e.g., feature fusion of RGB

TABLE 4 Indicators results of the six models on the test set.

Test indicators	Accuracy of spikelets (%)	Accuracy of florets (%)	F1-score(%)	mAP_0.5 (%)
YOLOv3	61.0	28.0	54.0	45.5
YOLOv7	82.4	62.2	71.5	72.5
Faster R-CNN	76.8	71	67.2	57.3
Cascade R-CNN	71.0	57.0	66.4	56.7
Based on CCF and superpixel segmentation	76.3	70.6	/	/
Improved YOLOv5s	**96.8**	**88.9**	**90.5**	**92.9**

images with the corresponding CCF; described in subsection 2.3.1) and some structural changes to the network (e.g., adding the Convolution block attention module, and the integrated Transformer small-target detection head; described in subsections 2.3.2 and 2.3.3). Therefore, an ablation experiment was conducted on the improved YOLOv5s model to explore the contribution of the proposed improvement strategies to the model detection performance improvement. The corresponding detection indicators for each optimization strategy are presented in Table 5.

As shown in Table 5, the improved YOLOv5s model proposed greatly enhanced various metrics of flowering detection in field wheat relative to the original model. The precision of the original YOLOv5s model was 71.9%, recall rate of 63.7%, F1-score of 67.6%, and mAP at 50% confidence of 70.5%. The P, R, F1-score, and mAP of the improved YOLOv5s model were increased by 23.4%, 22.5%, 22.9%, and 22.4%, respectively compared with the original YOLOv5s model. The *mAP* comparison test of the YOLOv5s prediction model based on different improvement strategies is shown in Figure 13.

Among the three optimization strategies, the detection model with the addition of feature fusion method improved the *P*, *R*, *F1-score*, and *mAP* by 12.6%, 7.3%, 9.6%, and 8.3%, respectively, compared with the original model. The effect of the detection model integrating the feature fusion method is shown in Figure 14A. It can be observed in the comparative test that the improved feature fusion method greatly improves the accuracy of the model, among which the recognition accuracy of florets and spikelets was increased by 5.5% and 10.8%, respectively. This is mainly because the feature fusion method reflects more texture characteristics of the wheat field images. As the target of florets is smaller than spikelets and the image background in RGB images is more complex, making it is difficult to extract the morphological features of florets. In addition, light points at the tip of spikelets and leaves could also be misidentified as florets under the influence of light, resulting in weak floret detection performance of the original YOLOv5s model. Thus, the performance of model detection was greatly improved by combining the feature fusion method.

The channel attention module and space attention module were introduced on the basis of the feature fusion method, which also improved the detection effect of the model. The precision, recall rate, *F1-score*, and *mAP* of the improved model were improved by 5.5%, 6.6%, 6.1%, and 8.1% respectively in the validation set. Experimental results show that the performance of the YOLOv5s model embedded with the convolutional block attention module was greatly improved.

As presented in Figure 14B the detection accuracy of florets and spikelets increased by 15.0% and 1.4% respectively, proving the effectiveness of the improvement. High-density planting and characteristic mapping of different traits are the main reasons for model performance degradation, on complex field images. After embedding the CBAM, the influence of occlusion and noise on the model can be weakened by assigning weights to different feature maps, and the useful target objects can be focused. Therefore, the improvement effect was significant.

The range of receptive fields obtained by different size detection heads also varies greatly, which reflects the ratio of feature maps to the input image area. When the receptive field is small, the number of elements in the original image is also small, thereby weakening the detection effect of larger targets. Conversely, when the receptive field is too large, the fine-grained information such as the spatial structure of small targets will be lost, leading to poor recognition effect of distant targets. In order to improve the detection performance of the model for small targets, the transformer detection head structure was added for tiny object detection. Combined with the other original three detection heads of YOLOv5s, the four-head structure can alleviate the negative influence caused by drastic changes in object size. After adding the integrated Transformer small-target detection head, the smallest detection box contained image size of 4x4 pixel, which solved the problem of high detection rate of small targets in wheat group images. The detection head structure added based on the above two types

TABLE 5 Detection indexes of the model with different optimization strategies.

	Flowering detection of wheat in field			
Feature fusion	×	√	√	√
Convolutional block attention model	×	×	√	√
Multi-detection heads structure integrated transformer	×	×	×	√
P (%)	71.9	84.5	90.0	**95.3**
R (%)	63.7	71.0	77.6	**86.2**
F1-score (%)	67.6	77.2	83.3	**90.5**
mAP_0.5 (%)	70.5	78.8	86.9	**92.9**

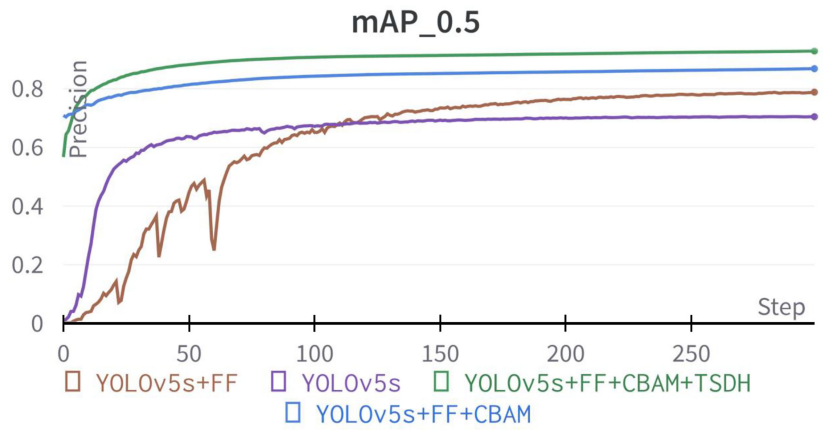


FIGURE 13
The mean average precision comparison of YOLOv5s prediction model based on different improvement strategies.

of improvement strategies improved the precision, recall rate, *F1-score*, and *mAP* of the model by 5.3%, 8.6%, 7.2%, and 6.0% respectively. As shown in Figure 14C, compared with the P-R curves of the model before adding the integrated Transformer small-target detection head, the improved model improved the recognition accuracy of florets by 7.7% and spikelets by 4.3%. The main improvement point is to optimize the recognition accuracy of the model for long-distance florets and spikelets and solve the problem of missing detection of long-distance tiny objects. After adding the integrated Transformer small-target detection head, although the performance index is less improved, it is essential to realize the accurate identification of florets and spikelets in field population images of wheat.

3.5 Comparisons using different attention methods

In this study, CBAM was integrated into the neck of YOLOv5 to improve the model. To evaluate the effectiveness of the CBAM, several state-of-the-art attention methods were applied to the improved YOLOv5s model for comparison. The selected attention methods include Squeeze-and-Excitation Networks(SENet) (Hu et al., 2018), Efficient Channel Attention(ECA-Net) (Wang Q. et al., 2020), Normalization-based Attention Module(NAM) (Liu et al., 2022), Coordinate Attention(CA) (Hou et al., 2021) and Effective Squeeze-Excitation(eSE) (Lee and Park, 2020). The evaluation metrics include *P*, *R*, *F1-score*, and *mAP*. The

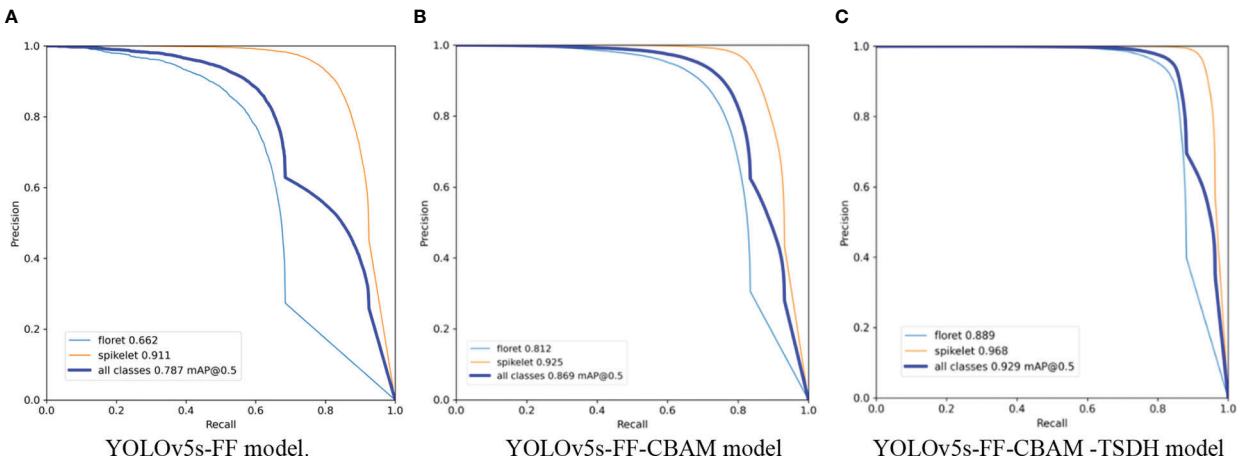


FIGURE 14
The precision-recall curves of florets and spikelets by different models. **(A)** precision-recall curve of YOLOv5s-FF model(floret 0.662 mAP@0.5, spikelet 0.911 mAP@0.5). **(B)** precision-recall curve of YOLOv5s-FF-CBAM model(floret 0.812 mAP@0.5, spikelet 0.925 mAP@0.5). **(A)** precision-recall curve of YOLOv5s-FF-CBAM-TSDH model (floret 0.889 mAP@0.5, spikelet 0.968 mAP@0.5).

TABLE 6 Comparisons of different attention methods under the improved YOLOv5s.

Model	P (%)	R (%)	F1-score (%)	mAP_0.5 (%)
YOLOv5s+FF+CBAM+TSDH	**95.3**	**86.2**	**90.5**	**92.9**
YOLOv5s+FF+SE+TSDH	83.1	71.0	76.6	80.0
YOLOv5s+FF+ESE+TSDH	85.3	71.4	77.7	81.2
YOLOv5s+FF+ECA+TSDH	84.2	71.6	77.2	81.0
YOLOv5s+FF+CA+TSDH	85.2	71.8	77.9	81.4
YOLOv5s+FF+NAM+TSDH	84.6	74.5	77.5	81.0

evaluation performance index results are shown in Table 6. Among the five attention mechanisms compared, the improved YOLOv5s model based on CA had the best comprehensive performance, the *P* of 85.2%, *R* of 71.8%, *F1-score* of 77.9%, and *mAP* of 81.4%. The improved model based on CBAM improved the P, R, F1-score, and mAP by 10.1%, 14.4%, 12.6%, and 11.5%, respectively, compared with the CA. The high density of wheat population in the field led to the shielding between wheat spikes, leaves, wheat awn, and stalks. The CBAM started from two scopes, channel and spatial, and allocated attention to two dimensions simultaneously, which enhances the effect of attention mechanism on model performance. The experimental results show that the performance of the improved YOLOv5s model based on CBAM had been greatly improved.

3.6 Flowering period experiment in field

The method proposed in this paper is based on the improved YOLOv5s model to achieve accurate identification of florets and spikelets, and to determine the flowering period of wheat based on the ratio of florets to spikelets. Hence, for the spikes of closed-flowering types, the presented method more likely will not work, because the key morphological features for recognition of florets (anthers) are enclosed. Therefore, the proposed method is suitable for wheat cultivars with open-flowering types of spikes, in which the stamens dangle from the florets.

Qualitative experiments showed that the improved model had good adaptability to the images of different angles, densities, and distances collected in the field. Since the camera can collect phenotypic information by video or timing shooting, this section will conduct the flowering period experiment in field to detect the image data obtained by different acquisition methods under different flowering conditions. The flowering rate was calculated based on the number of detected florets and spikelets to estimate the overall flowering situation of the plot and determine the flowering period. And the collection method with the highest reliability was selected for subsequent detection by comparing with the flowering rate measured manually. The artificial measurement of the flowering rate was calculated by the five-point sampling method on the flowering of each wheat spike in the plot. The field experimental process is displayed in Figure 15. The area of each field plot was 1.2 × 1.2 m, and it took about 12s for the acquisition platform to obtain data from a single



FIGURE 15
Field experiment process.

plot. During the field experiment, the camera was mounted on the side of the acquisition platform and tilted 45 degrees to obtain wheat data. When the Angle of overhead shooting is too large, the morphological structure of the distant spikelets and florets will be too small to be recognized, resulting in decreased accuracy. Due to this, the data obtained was partial plot images in the field. The three data acquisition methods of 0.5s interval shooting, 1s interval shooting, and video acquisition were used to determine the flowering period for field wheat. Due to the different time intervals of image data acquisition, the count of florets and spikelets identified by the improved YOLOv5s model was different under different acquisition methods. About 25 images were acquired in a single plot using the 0.5s interval shooting method, and about 13 images were acquired in a single plot using the 1s interval shooting method. The spikelets and florets detected under the three acquisition methods were superimposed respectively, and the determination of the wheat flowering period was performed based on the method mentioned in subsection 2.3. There was overlap between two adjacent image frames obtained. Repeated counting was used to reduce the influence of objective factors such as wind, wheat leaf occlusion, and the vibration of the acquisition platform during field data collection, which would lead to the failure of detection of some targets and decrease the accuracy. The comparison results of field experiments under different acquisition methods are shown in Table 7. Therefore, the least number of florets and spikelets were detected by the improved YOLOv5s model using the 1s interval shooting method compared with the others.

Under different flowering conditions, the flowering rates calculated manually of wheat and calculated by the improved YOLOv5s model under different acquisition methods were compared and analyzed. When the 0.5s interval shooting method was used for data acquisition, the error between the calculated flowering rate of the improved YOLOv5s model and the actual flowering rate was 2.4%, 4.1%, 3.2%, and 1.4% as shown in Table 7, respectively. It can be inferred that the 0.5s interval shooting method has higher reliability than the other two methods and it is used for subsequent detection. When the actual flowering rate exceeds about 30%, the error was negatively correlated with the actual flowering rate. When the actual flowering rate was about 0 to 30%, the error was positively correlated with the actual flowering rate. Inspection by analysis result revealed that the order of wheat flowering was from middle to top, and finally bottom. The middle spikelets develop faster, bloom, and pollinate first, and the anther size is similar to the bottom anther but slightly larger than the top anther. At flowering rates approaching 30%, there will be more apical florets of small size and not obvious, which is easy to cause

TABLE 7 The flowering period determination under different collection methods.

Collection method	Manually calculate flowering rate/%	Algorithm statistical results				Calculation error (%)
		Floret count	Spikelet count	Flowering rate (%)	Flowering period determination	
Video acquisition	9.3	2214	17407	12.7	heading period	3.4
0.5s interval shooting		87	1261	6.9	heading period	**2.4**
1s interval shooting		28	592	4.7	heading period	4.6
Video acquisition	32.4	7198	17951	40.1	heading period	7.7
0.5s interval shooting		351	1240	28.3	heading period	**4.1**
1s interval shooting		175	676	25.9	heading period	6.5
Video acquisition	57.7	15498	23482	66.0	flowering period	8.3
0.5s interval shooting		891	1635	54.5	flowering period	**3.2**
1s interval shooting		448	843	53.1	flowering period	4.6
Video acquisition	87.2	19813	20489	96.7	flowering period	9.5
0.5s interval shooting		1265	1474	85.8	flowering period	**1.4**
1s interval shooting		665	796	83.5	flowering period	3.7

(The manually calculated flowering rate as the actual flowering rate).

florets to be missed identification, so the error increases. The characteristics of florets become more obvious with the increase in flowering rate, and the missed detection rate also decreases. Therefore, the error is negatively correlated with the flowering rate when the flowering rate exceeds about 30%.

The flowering rate calculated by video detection is higher than that calculated manually in the three collection methods. The dynamic characteristics of floret changed little while that of spikelets changed strongly with the operation of the phenotypic platform in the field by analyzing the detection video frame-by-frame. In images with several frames apart, the continuous capture of florets was better, and spikelets were missed, so the calculated flowering rate was higher than that calculated manually. Compared with the method with the smallest error of 0.5s interval shooting, the error change of the detection results of the method with the 1s interval shooting is similar to that of the former. However, the prediction results were more likely to be affected by accidental factors due to the lower collection frequency, so the error is slightly greater than the detection method collected once every 0.5s interval.

4 Conclusion

The flowering period of wheat is one of the key agronomically valuable traits. To realize the real-time determination of the flowering period of wheat images in the field, a determination method based on the improved YOLOv5s model was proposed. Finally, the accurate detection of florets and spikelets was achieved, and the real-time determination of the wheat flowering period was completed based on the ratio of florets and spikelets. By fusing FF, CBAM, and TSDH, the improved YOLOv5s model attenuated the distortion of the extracted feature images caused by light, and solve the problem of obscured florets, spikelets missed detection, and difficult to be detected small targets in population images.

The proposed improved YOLOv5s model improved the accuracy of floret and spikelet recognition, with accuracy of 88.9% and 96.8%, respectively. The average detection time of a single image was 11.5ms. The average detection accuracy was higher than 86.4% under the complex conditions of strong light, drastic angular distortion, blurred target area, serious occlusion, and abundant population phenotype. The effectiveness and adaptability of the model in a variety of complex situations were proved. By conducting an ablation experiment to investigate the contribution of various strategies to model improvement. Among them, the Feature fusion method, the CBAM, and the integrated transformer multi-detection header structure showed an improvement in the mean average accuracy of the model by 8.3%, 8.1%, and 6.0%, respectively. Finally, the field experiment was carried out and the overall flowering rate of the plot was estimated based on the proportion of florets to spikelets. Compared with the results of artificial measurement of the flowering rate, the error of the model was less than 5% compared with the actual detection of the flowering rate, and the determination accuracy of the flowering period reached 100%, which meets the demand of practical application and demonstrates the feasibility of the research.

Our future work will focus on how to improve the detection speed of the model on the basis of ensuring the detection performance. At the same time, an efficient and accurate method for the whole growth period of wheat will be studied and applied to the breeding process, so as to provide data support for the improvement of the stable yield of wheat.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

PL, XS, and LL planned and designed the research, and wrote the manuscript. PL and XL advised on the design of the model and revised the manuscript. PL, XL, and CW supervised and revised the manuscript. WZ, YL, and JZ performed experiments partially. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by NSFC (31871543), Natural Science Foundation of Shandong Province(ZR2020KF002), and the project of Shandong Provincial Key Research and Development Plan (Major Science and Technology Innovation Project) (2021LZGC013; 2021TZXD001). The authors are grateful to all study participants.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ajlouni, M., Kruse, A., Condori-Apfata, J. A., Valderrama Valencia, M., Hoagland, C., Yang, Y., et al. (2020). Growth analysis of wheat using machine vision: opportunities and challenges. *Sensors* 20 (22), 6501. doi: 10.3390/s20226501

Bai, D., Sun, Y., Tao, B., Tong, X., Xu, M., Jiang, G., et al. (20222022). Improved single shot multibox detector target detection method based on deep feature fusion. *Concurrency Computation: Pract. Exp.* 34 (4), e6614. doi: 10.1002/CPE.6614

Bommert, P., Satoh-Nagasawa, N., Jackson, D., and Hirano, H.-Y. (2005). Genetics and evolution of inflorescence and flower development in grasses. *Plant Cell Physiol.* 46 (1), 69–78. doi: 10.1093/pcp/pci504

Cai, E., Baireddy, S., Yang, C., Delp, E. J., and Crawford, M. (2021). "Panicle counting in uav images for estimating flowering time in sorghum," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. 6280–6283, IEEE. doi: 10.1109/IGARSS47720.2021.9554291

Chen, X., Chen, Z., Xu, D., Lyu, Y., Li, Y., Li, S., et al. (2021). *De novo* design of G protein-coupled receptor 40 peptide agonists for type 2 diabetes mellitus based on artificial intelligence and site-directed mutagenesis. *Front. Bioeng Biotechnol.* 9. doi: 10.3389/fbioe.2021.694100

Cheng, X., Qiu, G., Hu, J., Wu, W., Chen, Y., Zhao, Y., et al. (2016). Control efficiency of different pesticides and application technologies against wheat scab. *J. Agricult.* 6 (1), 28–32. doi: 10.3969/j.issn.1007-7774.2016.01.006

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). "An image is worth 16x16 words: Transformers for image recognition at scale," in *arXiv preprint arXiv*, 2010 11929. doi: 10.48550/arXiv.2010.11929

Feng, Q., Huang, L., Sun, Y., Tong, X., Liu, X., Xie, Y., et al. (2022). Substation instrumentation target detection based on multi-scale feature fusion. *Concurrency Computation: Pract. Exp.* 34 (23), e7177. doi: 10.1002/CPE.7177

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13713–13722. doi: 10.1109/CVPR46437.2021.01350

Huang, L., Chen, C., Yun, J., Sun, Y., Tian, J., Hao, Z., et al. (2022). Multi-scale feature fusion convolutional neural network for indoor small target detection. *Front. Neurorobitics* 16. doi: 10.3389/fnbot.2022.881021

Huang, L., Fu, Q., He, M., Jiang, D., and Hao, Z. (20212021). Detection algorithm of safety helmet wearing based on deep learning. *Concurrency Computation: Pract. Exp.* 33 (13), e6234. doi: 10.1002/cpe.6234

Hu, W., Fu, X., Chen, F., and Yang, W. (2019). A path to next generation of plant phenomics. *Chin. Bull. Bot.* 54 (05), 558–568. doi: 10.11983/CBB19141

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. *Proc. IEEE Conf. Comput. Vision Pattern recognition*, 42 (8), 2011–2023. doi: 10.1109/CVPR.2018.00745

Lee, Y., and Park, J. (2020). "Centermask: Real-time anchor-free instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13906–13915. doi: 10.1109/CVPR42600.2020.01392

Li, H. Y., Chen, H. Y., Wang, L., Song, S. J., You, Z. H., Yan, X., et al. (2021). A structural deep network embedding model for predicting associations between miRNA and disease based on molecular association network. *Sci. Rep.* 11 (1), 12640. doi: 10.1038/s41598-021-91991-w

Liu, N., Li, H., Dou, Y., and Han, D. (2015). Morphology and structure analyses on inflorescence, spikelet and floret of bread wheat and its relatives. *J. Triticeae Crops* 35 (03), 293–299. doi: 10.7606/j.issn.1009-1041.2015.03.01

Liu, P., Liu, L., Wang, C., Zhu, Y., Wang, H., and Li, X. (2022). Determination method of field wheat flowering period baesd on machine vision. *Trans. Chin. Soc. Agric. Machinery* 53 (03), 251–258. doi: 10.6041/j.issn.1000-1298.2022.03.026

Liu, Y., Shao, Z., Teng, Y., and Hoffmann, N. (2021). NAM: Normalization-based attention module. *arXiv preprint arXiv*. doi: 10.48550/arXiv.2111.12419

Liu, C., Wang, K., Lu, H., and Zhiguo, C. (2021). 'Dynamic color transform for wheat head detection'. *Proc. IEEE/CVF Int. Conf. Comput. Vision*, 1278–1283. doi: 10.1109/ICCVW54120.2021.00148

Loshchilov, I., and Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv*. doi: 10.48550/arXiv.1608.03983

Ma, J., Li, Y., Du, K., Zheng, F., Zhang, L., Gong, Z., et al. (2020). Segmenting ears of winter wheat at flowering stage using digital images and deep learning. *Comput. Electron. Agric.* 168, 105159. doi: 10.1016/j.compag.2019.105159

Milicevic, M., Zubrinic, K., Grbavac, I., and Obradovic, I. (2020). Application of deep learning architectures for accurate detection of olive tree flowering phenophase. *Remote Sens.* 12 (13), 2120. doi: 10.3390/rs12132120

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788. doi: 10.1109/CVPR.2016.91

Sadeghi-Tehran, P., Sabermanesh, K., Virlet, N., and Hawkesford, M. J. (2017). Automated method to determine two critical growth stages of wheat: Heading and flowering. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00252

Sishodia, R. P., Ray, R. L., and Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote Sens.* 12 (19), 3136. doi: 10.3390/rs12193136

Velumani, K., Madec, S., De Solan, B., Lopez-Lozano, R., Gillet, J., Labrosse, J., et al. (2020). An automatic method based on daily *in situ* images and deep learning to date wheat heading stage. *Field Crops Res.* 252, 107793. doi: 10.1016/j.fcr.2020.107793

Wang, K., Dong, S., Li, F., Gun, J., Tai, S., Wang, L., et al. (2020). QTL mapping and analysis of heading time and flowering time of wheat. *Shandong Agric. Sci.* 2020 52 (01), 17–23. doi: 10.14083/j.issn.1001-4942.2020.01.003

Wang, K., Dong, S., Li, F., Guo, J., TAI, S., Wang, L., et al. (2020). And analysis of heading time and flowering time of wheat. *Shangdong Agric. Sci.* 52 (01), 17–23. doi: 10.14083/j.issn.1001-4942.2020.01.003

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "Eca-net: Efficient channel attention for deep convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE*. 11531–11539. doi: 10.1109/CVPR42600.2020.01155

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 3–19. doi: 10.48550/arXiv.1807.06521

Xiong, J., Liu, B., Zhong, Z., Chen, S., and Zheng, Z. (2021). Litchi flower and leaf segmentation and recognition based on deep semantic segmentation. *Trans. Chin. Soc.* 52 (06), 252–258. doi: 10.6041/j.issn.1000-1298.2021.06.026

Yun, J., Jiang, D., Liu, Y., Sun, Y., Tao, B., Kong, J., et al. (20222022). Real-time target detection method based on lightweight convolutional neural network. *Front. Bioengineering Biotechnol.* 10. doi: 10.3389/fbioe.2022

Zhang, H., Zhou, H., Zheng, J., Ge, Y., and Li, Y. (2020). Research progress and prospect in plant phenotyping platform and image analysis technology. *Trans. Chin. Soc. Agric. Machinery* 51 (03), 1–17. doi: 10.6041/j.issn.1000-1298.2020.03.001

Zhao, W., LIi, S., Li, A., Zhang, B., and Chen, J. (2021). Deep fusion of hyperspectral images and multi-source remote sensing data for classification with convolutional neural network. *Natl. Remote Sens. Bull.* 25 (07), 1489–1502. doi: 10.11834/jrs.20219117

Zhao, Z., Tong, X., Sun, Y., Bai, D., Liu, X., Zhao, G., et al. (2022). "Large Scale instance segmentation of outdoor environment based on improved YOLACT," in *Concurrency and computation: Practice and experience*. 34 (28), e7370. doi: 10.1002/CPE.7370

Zhao, C., Wen, C., Lin, S., Guo, W., and Long, J. (2020). Tomato florescence recognition and detection method based on cascaded neural network. *Trans. Chin. Soc. Agric. Eng.* 36 (24), 143–152. doi: 10.11975/j.issn.1002-6819.2020.24.017

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop phenomics: Current status and perspectives. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00714

Zhu, R., Li, S., Sun, Y., Cao, Y., Sun, K., Guo, Y., et al. (2021). Research advances and prospects of crop 3D reconstruction technology. *Smart Agric.* 3 (03), 94–115. doi: 10.12133/j.smartag.2021.3.3.202102-SA002

Zhu, X., Lyu, S., Wang, X., and Zhao, Q. (2021). 'TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios'. *Proc. IEEE/CVF Int. Conf. Comput. Vision*, 2778–2788. doi: 10.1109/ICCVW54120.2021.00312

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 13 January 2023
DOI 10.3389/fpls.2022.1016507

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Kai Jia,
Massachusetts Institute of Technology,
United States
John Doonan,
Aberystwyth University,
United Kingdom

*CORRESPONDENCE
Yuzuko Utsumi
yuzuko@omu.ac.jp

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 11 August 2022
ACCEPTED 19 December 2022
PUBLISHED 13 January 2023

CITATION
Kinose R, Utsumi Y, Iwamura M and Kise K (2023) Tiller estimation method using deep neural networks. *Front. Plant Sci.* 13:1016507. doi: 10.3389/fpls.2022.1016507

COPYRIGHT
© 2023 Kinose, Utsumi, Iwamura and Kise. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Tiller estimation method using deep neural networks

Rikuya Kinose[1], Yuzuko Utsumi[1,2]*, Masakazu Iwamura[1,2] and Koichi Kise[1,2]

[1]Graduate School of Engineering, Osaka Prefecture University, Sakai, Japan, [2]Graduate School of Informatics, Osaka Metropolitan University, Sakai, Japan

This paper describes a method based on a deep neural network (DNN) for estimating the number of tillers on a plant. A tiller is a branch on a grass plant, and the number of tillers is one of the most important determinants of yield. Traditionally, the tiller number is usually counted by hand, and so an automated approach is necessary for high-throughput phenotyping. Conventional methods use heuristic features to estimate the tiller number. Based on the successful application of DNNs in the field of computer vision, the use of DNN-based features instead of heuristic features is expected to improve the estimation accuracy. However, as DNNs generally require large volumes of data for training, it is difficult to apply them to estimation problems for which large training datasets are unavailable. In this paper, we use two strategies to overcome the problem of insufficient training data: the use of a pretrained DNN model and the use of pretext tasks for learning the feature representation. We extract features using the resulting DNNs and estimate the tiller numbers through a regression technique. We conducted experiments using side-view whole plant images taken with plan backgroud. The experimental results show that the proposed methods using a pretrained model and specific pretext tasks achieve better performance than the conventional method.

KEYWORDS

tiller number estimation, deep neural network (DNN), pretext task, self-supervised learning, regression

1 Introduction

A tiller is a branch of a grass plant. For grain bearing members of the grass family, the number of fertile shoots per unit area, number of grains per ear, and size of grains are the determinants of yield. Therefore tillering is one of the traits targeted for phenotyping, particularly as tiller number can vary through the life of a plant in response to environmental and genetic factors (Xie et al., 2015). Therefore, it is one of the traits that is targeted for phenotyping. Destructive surveys have commonly been used to count the number of tillers, because they are hard to count visually; leaves and tillers look similar, and the density of tillers tends to be highest at the base of the plant. However,

destructive surveys present a bottleneck to phenotyping tasks because they are time-consuming and labor-intensive, making it impossible to trace the growth of the plants. To achieve nondestructive and automatic tiller number estimation, several image-based methods have been proposed (Fahlgren et al., 2015; Boyle et al., 2016).

However, the estimation accuracy of the conventional image-based methods is generally poor. However, the estimation accuracy of the conventional image-based methods is generally poor. To estimate the tiller number, image-based approaches use hand-crafted features[1] such as the area and aspect ratio of a plant within an image and the output of the Frangi filter (Frangi et al., 1998) for linear regression. As these methods only use a few heuristic features of the plants' appearance, they do not take full advantage of the information contained in the images. The recent development of image recognition techniques using features learned by deep neural networks (DNNs) surpasses the performance of conventional hand-crafted feature-based methods (Taigman et al., 2014; Simonyan and Zisserman, 2015; Schroff et al., 2015; Hu et al., 2018). DNNs learn image features directly from the image appearance. Thus, the features learned by DNNs take full advantage of the plants' appearance. This motivates us to use DNNs to learn features as a means of realizing high-accuracy tiller number estimation.

DNNs requires large volumes of training data, consisting of pairs of an image and the corresponding ground truth. The image dataset of Setaria plants (Gehan et al., 2015) contains only around 600 images with the corresponding tiller numbers because the operation of counting the tiller numbers is time-consuming and labor-intensive, as mentioned above. Therefore, it is difficult to prepare sufficient training data for DNNs, making it almost impossible to apply DNN-based methods for tiller number estimation.

As a lack of training data is commonly encountered in the field of computer vision and pattern recognition, several methods have been developed to enable DNNs to be used with small-scale data. For example, transfer learning (Huang et al., 2019) transfers the network learning to another dataset, semi-supervised learning (Miyato et al., 2019) uses partly labeled data for learning, and self-supervised learning (Gidaris et al., 2018) uses self-generating labels. Some self-supervised learning methods that learn features by solving other tasks have achieved comparable performance to supervised methods (Noroozi et al., 2017; Gidaris et al., 2018; Noroozi et al., 2018). These other tasks are called "pretext tasks," and they can be applied to problems in which large numbers of unlabeled data are available.

In this paper, we describe the use of self-supervised learning and transfer learning to estimate the tiller number, even though there are relatively few training data (Utsumi et al., 2019; Kinose et al., 2022). To the best of our knowledge, this is the first attempt to use deep learning-based image features in nondestructive tiller number estimation using a single RGB image. We apply transfer learning to the estimation task and examine how the features learned from other data affect the estimation. We also set some pretext tasks for learning DNNs and evaluate how the pretext tasks enhance the estimation performance. Experimental results show that the proposed method outperforms the conventional method and that the pretext tasks enhance the estimation accuracy. The results showed that when using the framework of the proposed method, the plant trait can be estimated accurately using deep-learning even though few training data are acquired.

1.1 Related work

1.1.1 Tiller number estimation

DNN-based tiller number estimation techniques have already been proposed (Deng et al., 2020; Wu et al., 2021). Deng et al. (Deng et al., 2020) applied DNN-based image detection to stubble images as a means of counting the tillers. However, this method requires a destructive survey, making it difficult to track the growth traits of the plants. The idea of counting tillers proposed by Wu et al. (Wu et al., 2021) is almost the same as that developed by Deng et al., except that the images are obtained using micro-CT. Unfortunately, micro-CT is too expensive to be widely used. Different from these methods, the proposed requires only an RGB image to estimate the tiller numbers. Therefore, it is suitable for easy and high-throughput phenotyping.

1.1.2 Image-based plant phenotyping using DNNs

The most common task for DNN-based individual phenotyping is leaf counting because an image dataset of Arabidopsis thaliana was released (Minervini et al., 2016). The dataset has since been used in the development of many methods (Aich and Stavness, 2017; Ubbens et al., 2018; Ward et al., 2018). However, the dataset has few image data in which the number of leaves is identified. Therefore, techniques that artificially increase the number of data using data synthesis based on plant models have been proposed, enabling DNNs to be applied to small sets of labeled data (Ubbens et al., 2018; Ward et al., 2018). This data synthesis approach cannot be easily applied to tiller number estimation because the structure of grass plants is too complicated to model.

In addition to counting the leaves of Arabidopsis thaliana, many traits have been estimated using DNNs. Roots are another

1 In computer vision and pattern recognition area, image features which are determined by reserachers are called "hand-crafterd features," contrast to the deep-learning-based features, which are automatically determined by training.

typical subject for trait estimation using DNN-based image analysis. For example, segmentation algorithms for root regions (Han and Kuo, 2018; Wang et al., 2019; Gaggion et al., 2021) and root structure analysis based on the characterization of roots (Wu et al., 2018; Yasrab et al., 2019) have been proposed. Certain traits of wheat, which is a member of the grass plant family, have also been estimated, such as the number of spikes and spikelets (Pound et al., 2017) and the emergence and biomass (Aich et al., 2018).

1.1.3 Pretext tasks

Various pretext tasks have been proposed. For example, colorizing images (Zhang et al., 2016), solving jigsaw puzzles (Noroozi and Favaro, 2016), predicting image rotations (Gidaris et al., 2018), and counting the number of objects within an image (Noroozi et al., 2017) have been used for representation learning. The learned representations are used for image segmentation, image recognition, and object recognition.

In establishing the proposed method, we set some pretext tasks for tiller number estimation according to these previous methods. The application of pretext task means that tiller number estimation can be conducted using DNNs, even if few labeled data are available.

2 Materials and methods

We explain how the proposed method estimates the tiller number from an image. We adopt regression-based estimation for tiller counting, as in conventional image-based tiller number estimation methods (Fahlgren et al., 2015; Boyle et al., 2016). This is because regression-based estimation is more practical than the detection-based method. For examples, in the leaf counting task of Arabidopsis thaliana, regression-based method show better accuracy than the object-detection-based method Ubbens and Stavness (2017), and many regression-based method have been proposed Giuffrida et al. (2015); Dobrescu et al. (2017); Aich and Stavness (2017). Tillers have a similar appearance to leaves, and so it is hard to detect tillers from images. Moreover, the tillers become too dense to detect as the plant grows. Therefore, we adopt a regression-based method.

The proposed method consists of two parts: feature learning and estimating the tiller numbers. Figure 1 shows an overview of the feature learning part. The VGG-16, which is one of the most popular DNN models, is pretrained on the ImageNet classification task and pretext tasks. After pretraining, fully connected layers are discarded and new ones are prepared according to the tasks. Although it is common to use the ImageNet dataset for pretraining (Figure 1A), we use images without tiller number labels on the pretext tasks (Figures 1B–D). The labels on the pretext tasks were acquired automatically by image processing. Figure 2 shows an overview of the tiller number estimation part. In the tiller number estimation, the features are extracted by the trained networks, which purged the FC layers. The dimensionality of the feature is 4096. The tiller numbers are estimated by regression; the features tiller numbers are used as an independent and dependent value, respectively. A small number of images with tiller numbers are used for training the regression model of tiller number estimation. Image resources and processing, the pretrained model, pretext tasks, and regression models are now described in detail.

2.1 Image resources and processing

We used the dataset that appears in Gehan2015[2,3] The first row in Figure S1 shows some examples of the dataset. The dataset contains 25,570 images of potted Setaria taken from the side in a controlled laboratory environment. The species of the Setaria are S. viridis (A10), S. italica (B10), and eight RILs (RIL020, RIL070, RIL098, RIL102, RIL128, RIL133, RIL161, RIL187) from an S. viridis × S. italica population. We used side-view whole plant images in the dataset. The images are in RGB color, and the image resolution is 2,454 × 2, 056 pixels. In the dataset, 576 images have tiller numbers that were counted manually. Thus, there are 24,994 unlabeled images that have no tiller number. Many of the unlabeled images were taken at the same time as the labeled images. To avoid mixing unlabeled data that are similar to the labeled data, we only used the 22,110 unlabeled images that were not taken at the same time as the labeled images. There are some images in which plants stick out from the frame in the dataset.

We normalized the images before the experiments. The magnification of the images was artificially determined according to the plant growth degree. As the first row of Figure S1 shows, the pot size and the background differ depending on the plant size. If such images were used for learning, the network may learn features that focus on changes in the pots and backgrounds. To avoid the network focusing on parts unrelated to the plants, we normalized the images. We executed the normalization in a semi-automatic manner: We sampled an image from each magnification and manually cropped a rectangle area that included the whole plant area to remove the background. Because all plants were in pots of the same size, the images were resized so that the pot size was the same. After removing the background, we manually determinated the upper part of the pot in the cropped image. The cropped image was translated to place the center of the pot in the center of the image and resized the cropped image because

2 https://doi.org/10.6084/m9.figshare.1272859.v12

3 https://plantcv.danforthcenter.org/pages/data-sets/2013/setaria_burnin2.html.

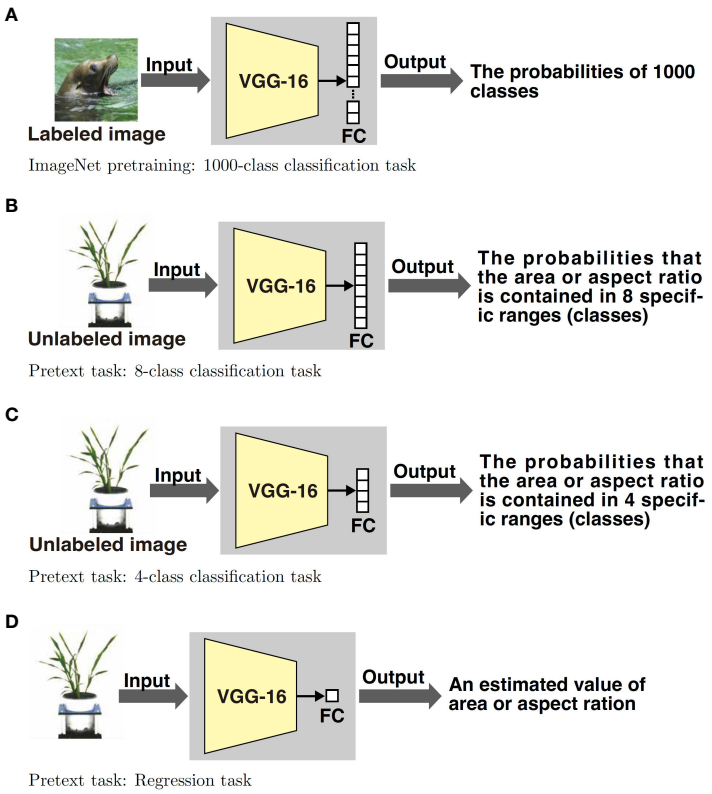


FIGURE 1
We use four feature training methods. Figure 1A shows a typical supervised training method, which uses the ImageNet dataset. The dataset consists of 1000 classes; thus the output layer (the last fully connected (FC) layer) of the network consists of 1000 elements. The DNN model is trained by updating the parameters in order to reduce the error between the output of the DNN and the ground truth. We use a pretraining model available on the web. Figures 1B–D show self-supervised methods using pretext tasks; they use 8-class classification, 4-class classification, and regression tasks. Hence, their last layers consist of eight, four, and one element, respectively. The pretext tasks in the proposed method estimate the area or aspect ratio of the plant within the input image. The images for training are unlabeled with the tiller numbers. The ground truth of the area and aspect ratio of the plant within the image are calculated by using image processing beforehand.

the pot size was 32 pixels. The images were padded with white pixels to make the images square for input to the network. Finally, the size of the image is 224 × 224 pixels. The rest of the images were automatically cropped in the same area, translated by the same amount, and scaled to the same size as the sample image. We confirmed that all plant areas were not cut off in the normalized images. All procedures were performed using OpenCV, and we used the bicubic method for pixel interpolation when the images were resized. There were no images that some parts of larger plants are out of the frame caused by the normalization.

We used the 22,110 unlabeled images for the training of the pretext tasks and the 576 images which has tiller numbers for evaluating tiller number estimation. The unlabeled images did not overlap with the labeled 576 images.

2.2 Feature extraction

2.2.1 Pretrained model

A sufficient number of labeled data for training is required for a DNN to achieve good performance. However, it is often the case that a sufficient number of labeled data are not available. A typical solution is to use a pretrained model. Usually, a pretrained model is trained on a large dataset, such as the ImageNet dataset (Deng et al., 2009), in a classification task. The use of the pretrained model is considered reasonable from the observation that a DNN trained on a large dataset in a task extracts effective features in a different task.

There are some standard DNNs used in the field of computer vision. One of such DNNs is the VGG model (Simonyan and Zisserman, 2015). Compared to ResNet He et al. (2016), another

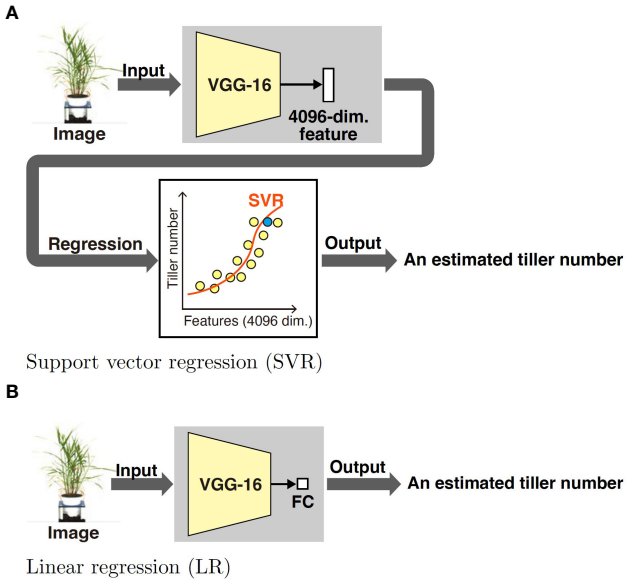


FIGURE 2
We propose two tiller number estimation methods using different regression models: support vector regression (SVR) and linear regression (LR). Both methods use the pretrained DNN model in Figure 1; more specifically, we use the pretrained the VGG-16 model. In both methods, a plant image taken from the side is used for estimating the number of tillers. In SVR, a 4096-dimensional feature is extracted from the input image using the pretrained DNN. The parameters of SVR are estimated using features extracted from labeled images; the features and tiller numbers are used as independent and dependent values, respectively. In LR, a new FC layer consisting of an element is prepared. Using labeled images, the DNN is trained. Then, the DNN outputs an estimated tiller number for an input unlabeled data.

standard DNN, GG is simple and easy to train. Although ResNet often achieves better performance in a complex task with a lot of labeled data for training, VGG often performs equivalently in a simple task with less labeled data. Since our task is simple, we use the 16-layer VGG model, which is called VGG-16.

2.2.2 Pretext tasks

As mentioned in Section 1, it is redimpractical to learn the feature expression from the tiller number estimation task directly because of the shortage of labeled training data. Thus, we use pretext tasks to learn the feature expression, and estimate the tiller number using the learned features.

The VGG-16 model (Simonyan and Zisserman, 2015) is trained using pretext tasks that predict appearance related values acquired automatically from a plant image. As shown in Figure 1, we set two pretext tasks: estimating the area of a plant within an image and estimating the aspect ratio of a plant. We consider the area and aspect ratio because they were used as the dependent variables for estimating the tiller numbers in a previous study (Fahlgren et al., 2015) and are expected to provide good feature expressions for tiller number estimation.

We investigate two methods of estimating the area or aspect ratio of the plant within an image in the pretext tasks: the area or aspect ratio themselves and the discretized area or aspect ratio. When estimating the area or aspect ratio itself, as shown in Figure 1D, the network is trained so that the output is the area or aspect ratio. We call the pretext task that estimates the value itself "regression task," because in this case, the pretext task can be regarded as a regression task with the image as the independent value and the continuous values of area and aspect ratio as the dependent values. In the case of estimating discrete area or aspect ratio, instead of outputting a numerical area or aspect ratio, the network predicts the discretized area or aspect ratio of the plants in the input image, as shown in Figures 1B and C. Therefore, predicting discrete area or aspect ratio is equivalent to classification. We call the pretext task estimating the discretized values "classification task."

We conducted network training on the pretext tasks using the normalized images. The ground truth of the pretext tasks was calculated automatically using image processing. Following the "Single plant RGB image workflow" in the PlantCV tutorial[4], the normalized images were translated into HSV and Lab images, and thresholding was applied to the saturation component of the HSV images and the *a* and *b* components of the Lab images. The plant area was then segmented by taking the logical sum of the threshold results. The area and aspect ratio were calculated from the segmented plant area. All processes were conducted using PlantCV [5]. The images were divided into four or eight classes in the classification task according to the area and aspect ratio values, respectively. The images were divided so that the number of images in each class became the same.

4 https://plantcv.readthedocs.io/en/stable/tutorials/vis_tutorial/

5 https://plantcv.danforthcenter.org/

The network was trained to predict the class to which the input image belongs. In the regression task, the network was trained to predict the area or aspect ratio of the input images. Both tasks used 80% of the images for training and 20% of the images for testing. The network used for training was the VGG-16 model pretrained by the ImageNet dataset. The mini-batch size, learning rate, and epochs for the training were set as 128, 0.0001, and 200, respectively. We applied horizontal and vertical flip data augmentation. We trained the network 12 times with the above condition and adopted the model that gave the lowest training error for tiller number estimation. We used the Keras TensorFlow2 backend to execute the training process.

2.3 Tiller number estimation

We use two regression models to estimate the tiller numbers, namely support vector regression (SVR) and linear regression (LR).

SVR involves the application of a support vector machine to regression. The most significant advantage of SVR is that it deals with nonlinear regression problems through the same framework as linear SVR. In SVR, a feature space can be mapped to a space of much higher dimension using a kernel function. When the kernel function is nonlinear, SVR can deal with nonlinear regression problems. Moreover, SVR can learn from small-sized datasets. Hence, we apply SVR to tiller estimation. Specifically, we extract features from labeled images using the models described in Sections 2.2.1 and 2.2.2, and then apply SVR.

We also use linear regression (LR) for the estimation task. LR is one of the simplest regression methods and is equivalent to a fully connected neural network without a hidden layer. Because it is easy to implement LR with methods using DNN-based features, we apply LR for the estimation task. As with SVR, we learn the LR model using the features extracted from labeled images.

We estimated the tiller number using the features extracted by pretext-task-trained and ImageNet pretrained models. In the case of SVR, we used scikit-learn[6] for the implementation, which is one of the most popular machine learning libraries for Python. The radial basis function was used as the kernel. The cost parameter C and parameter ϵ were set to 100 and 1.0, respectively, and default values were used for the other parameters. LR was implemented by adding two fully connected layers to the VGG-16 model. We then trained only the added layer while freezing VGG-16.

6 https://scikit-learn.org/stable/

3 Results

3.1 Tiller number estimation

We executed the proposed tiller number estimation using features extracted by models trained by pretext tasks and the pretrained model to reveal the difference between the feature extraction models. We used six-fold cross-validation to calculate the accuracy of the tiller number estimation. That is, the images were divided into six groups and the regression models were trained with five groups and validated with the remaining group. This process was repeated until all groups had been used for validation. The accuracy of the model was calculated by taking the average of each of the six cross-validation tasks. We adopt the mean absolute error (MAE) to evaluate the accuracy of the proposed method. We used the GPU, NVIDIA TITAN RTX, for training the network with the pretext tasks, and the CPU server, which has Opteron 6348 CPU (2.8GHz) and 512GB memory for estimating the tiller numbers. We also executed the conventional method proposed in Fahlgren2015 on the same CPU server.

For fare comparison, we executed the method proposed by Fahlgren et al., 2015 with the dataset we used. Fahlgren et al., 2015 estimated plant fresh weight using plant area on an original image by the following equation.

$$M_{fw} = 3.755 \times 10^{-5} A_{sv} - 0.2704 \quad (1)$$

M_{fw}, A_{sv} are estimated plant fresh weight and area of the plant in an image, respectively. Then, the tiller number was estimated by using the estimated fresh weight and aspect ratio of the plant in the image as follows:

$$TC = 0.22 M_{fw} - 2.19 HW + 5.26, \quad (2)$$

TC,HW are the tiller number and aspect ratio of the plant in the image. We cannot directly apply the equations as we resized the original images. Therefore, we estimated the parameters in eq 1 using images with fresh weight. The original dataset we used in this paper(https://figshare.com/articles/dataset/DDPSC_Phenotyping_Manuscript_1_Files/1272859) has 158 images that have fresh weight. We normalized the images in the same manner as other images used for the experiments and estimated the parameter of eq. 1. We estimated the parameter 2, and evaluated the accuracy of the equation with the same 576 images, which have tiller numbers, as the proposed methods were evaluated using six-fold cross-validation. The estimated parameters of eq. 1 are as follows:

$$M_{fw} = 0.005790 A_{sv} - 0.3372. \quad (3)$$

The coefficient of determination of eq. 3 R^2 was 0.9589. The estimated parameters of eq. 2 are as follows:

$$TC = 0.1786 M_{fw} - 1.1102 HW + 3.885 \quad (4)$$

Table 1 presents the MAE when using SVR and LR to estimate the tiller numbers. The total running time for estimation per image was 98, 191, and 0.6 ms when using LR with the proposed method, SVR with the proposed method, and (Fahlgren et al., 2015), respectively. We also evaluated the standard error and 95% confidence interval of the estimation, as shown in Figure 3.

3.2 Individual estimation results

The measured tiller number (horizontal axis) and estimated tiller number (vertical axis) are compared in Figures 4, 5 for the cases using SVR and LR for tiller number estimation, respectively.

4 Discussion

This proposed method is the first attempt to apply self-supervised learning using pretext tasks for plant phenotyping, as far as we know. Plant datasets have insufficient labeled data for applying DNNs. The proposed semi-supervised method for estimating the number of tillers requires only a few labeled data. Therefore, the proposed method show good estimation accuracy. The best MAE of 0.57 is achieved when the area regression is used for the pretext task and LR is used to predict the tiller numbers. The MAE by Fahlgren et al. (Fahlgren et al., 2015) was 1.187. Note that it is not possible to make a general comparison because of the different image usage conditions and because Fahlgren et al. (Fahlgren et al., 2015) used a different number of images to that in the dataset (Gehan et al., 2015). However, it appears that the proposed method achieves good accuracy. The proposed method show good estimation accuracy. The best MAE of 0.57 is achieved when the area regression is used for the pretext task and LR is used to predict the tiller numbers. The MAE by Fahlgren et al. (Fahlgren et al., 2015) was 1.187. Note that it is not possible to make a general comparison because of the different image usage conditions and because Fahlgren et al. (Fahlgren et al., 2015) used a different number of images to that in the dataset (Gehan et al., 2015). However, it appears that the proposed method achieves good accuracy.

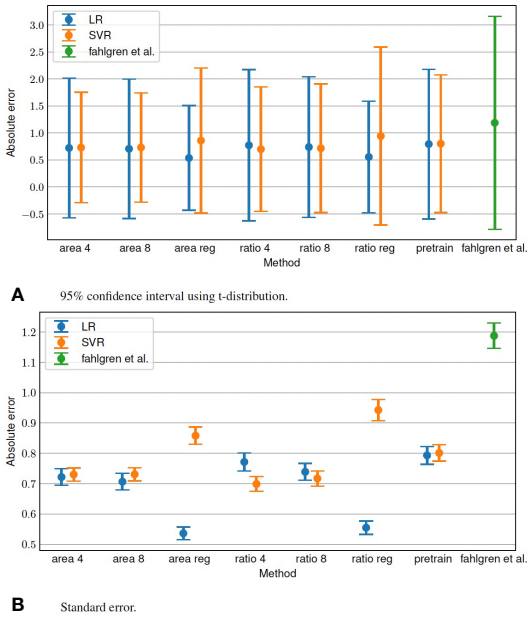


FIGURE 3
Standard error and 95% confidence interval of tiller number estimation.

To clarify the effect of feature learning in the pretext tasks, we compared the accuracy of the pretext tasks and pretrained models. Many of the pretext tasks resulted in higher accuracy than using the pretrained model. Therefore, learning features using pretext tasks contributes to improving the accuracy of estimating tiller numbers.

The tiller number estimation accuracy depends on the trait estimated in the pretext task. The tiller number estimation accuracy is better when the area is used in the pretext task than when the aspect ratio is used. Therefore, the features learned in the pretext task using the area are more effective for tiller estimation than those learned from the aspect ratio.

The pretext task that gives the better tiller number estimation accuracy also depends on the tiller number

TABLE 1 MAE of estimation results when using SVR and LR.

		Pretext task					
		Area			Aspect ratio		
Reg. model	Pretrained	4 cls.	8 cls.	Reg.	4 cls.	8 cls.	Reg.
(r)1-1(r)2-8 SVR	0.80	0.74	0.78	0.91	0.73	0.73	1.00
LR	0.79	0.74	0.71	0.57	0.96	1.06	0.62

"Reg. model" stands for the regression model used for tiller number estimation. "Reg.", "4 cls.", "8 cls." in the pretext task stands for Regression task, 4 class classification task, and 8 class classification task, respectively.

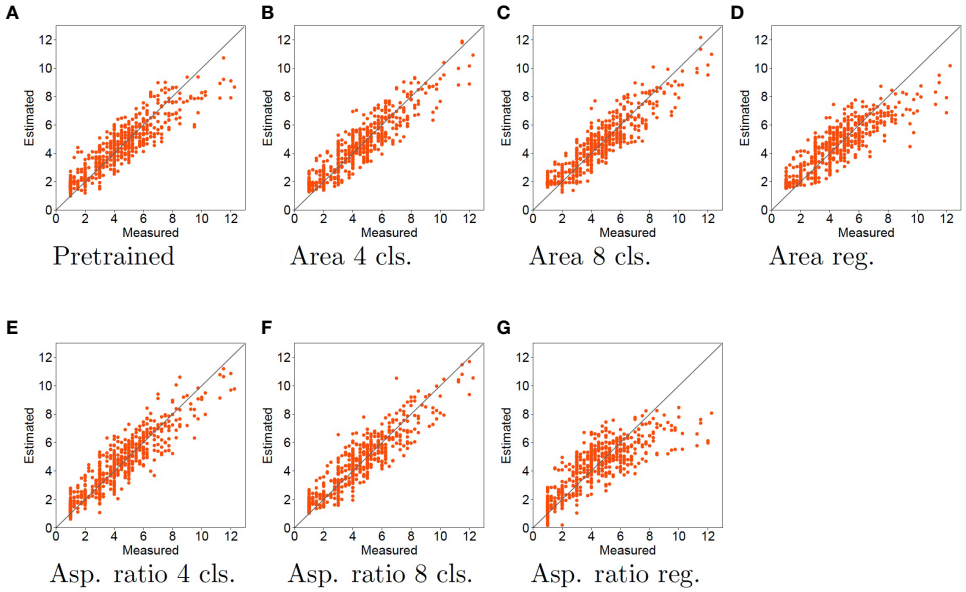


FIGURE 4
Experimental results using SVR for tiller number estimation. Horizontal and vertical axes represent the measured and estimated tiller numbers, respectively. Each red dot denotes a sample for the estimated tiller number. The black line indicates the case where the measured and estimated data match. Therefore, the closer the points are to the black line, the more accurate is the estimation.

estimation method. When SVR is used, the application of classification in the pretext task results in better accuracy than regression. In contrast, when LR is used, the application of regression in the pretext task achieves better accuracy than classification.

There is clearly a different tendency when SVR and LR are used for tiller number estimation. As shown in Figure 4A, when the pretrained model features are used with SVR, the estimated tiller number is substantially underestimated when the measured tiller number is high. In Figures 4B, C, E, F, orange dots are

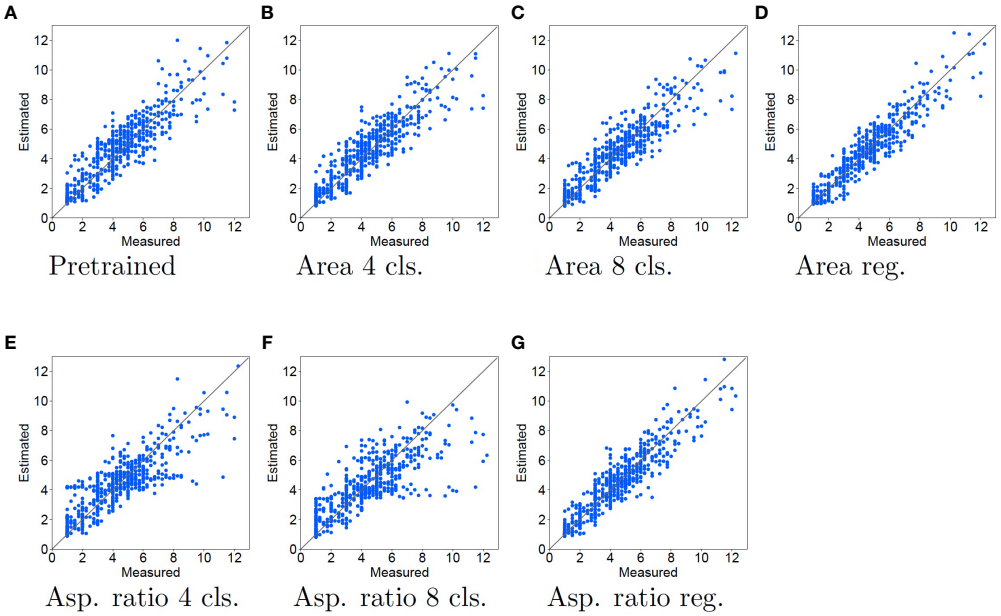


FIGURE 5
Experimental results using LR for tiller number estimation. The contents of the graph are the same as in Figure 4.

distributed close to the diagonal lines. This means that the accuracy of the tiller number estimation improves for samples with larger tiller numbers when the features learned by classification tasks are used, compared with pretrained model features. Thus, using classification for the pretext tasks improves the tiller number estimation accuracy. However, when using regression for the pretext tasks, the estimation accuracies are worse than those with the pretrained model. In particular, as shown in Figures 4D, G, orange dots are plotted in the rightmost of the figures [approximately 9 to 12 of measured (horizontal axis)],. This means that the estimation results for samples with larger tiller numbers are worse than those using the pretrained model.

When regression and LR were used for the pretext task and tiller number estimation, respectively, dots are close to the diagonal line, as shown in Figures 5D, G. This means that the estimation accuracy improves for all samples. In particular, comparing the pretrained model with the regression pretext tasks, the top right dots of regression pretext tasks is more close to the diagonal lines. This means that the accuracy is enhanced for samples with large measured tiller numbers. In contrast, as shown in Figures 5B, C, E, F, dots are vary widely from the diagonal lines. This means that when the features learned by the classification task are used, the estimation accuracy is the same or worse than that of the pretrained model. When the aspect ratio is used for the classification task, the estimation accuracy becomes worse, with the estimated tiller numbers consistently lower than the measured values.

We used images which are taken well-controlled lab environment and taken separately. Therefore, the proposed method would work well on images taken in a similar environment but not on images taken in a different environment. For example, if the images were taken in the field, multiple plants would appear in the images. In this case, we need to recognize the individual plants and apply the proposed method to each plant. However, when the plans are crowded, recognizing individual plants in side-view images is difficult for current image recognition. Thus, the proposed method is hard to apply to the images taken in the field.

However, some improvements would make the proposed method applicable to the images taken in some different environments. When the images were taken under different lighting conditions, we can apply the proposed method to the images by adding the different lighting condition images for training the network and the regression model for estimating tiller numbers. When using the images taken with noisy backgrounds, we can apply the proposed method by using a plant detection method such as Amean et al., 2021 to delete the noisy background.

In future work, we will use other pretext tasks to learn the feature representations. The mechanisms of the pretext tasks remain obscure, and it is not known what kinds of pretext tasks are most effective for a given object task. Therefore, we will attempt to determine the most appropriate pretext task for the object task by trial and error. We also plan to apply the proposed method to other grass plant family such as wheat and rice.

Additionally, we will apply the proposed method to other plant phenotyping tasks. The proposed method assumes that few labeled training data are available. This is typically true of plant phenotyping tasks because many appearance traits are measured manually. We expect that the proposed method will be helpful in automating the measurement of various traits.

5 Conclusion

This paper has proposed a DNN-based tiller estimation method that achieves improved performance compared with conventional methods. The proposed method uses two separate models for feature extraction: a pretrained VGG-16 model and a model produced by solving pretext tasks. We considered both SVR and LR to estimate the tiller numbers. Experimental results show that the pretrained model and the model based on pretext tasks allow the proposed method to outperform the conventional approach.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

YU and MI contributed to the conception and design of the study. RK performed the statistical analysis. KK prepared the materials for the research. YU wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by JSPS KAKENHI Grant Number JP18K18074.

Acknowledgments

This paper is open as a preprint paper on Research Square kinose2022. We thank Stuart Jenkinson, PhD, from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1016507/full#supplementary-material

References

Aich, S., Josuttes, A., Ovsyannikov, I., Strueby, K., Ahmed, I., Duddu, H. S., et al. (2018). "DeepWheat: Estimating phenotypic traits from crop images with deep learning," in *Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. (Lake Tahoe, NV, USA: IEEE), 323–332. doi: 10.1109/WACV.2018.00042

Aich, S., and Stavness, I. (2017). "Leaf counting with deep convolutional and deconvolutional networks," in *Proceedings of ICCV 2017 Workshop on Computer Vision Problems in Plant Phenotyping*. (Venice, Italy: IEEE), 2080–2089. doi: 10.1109/ICCVW.2017.244

Amean, Z. M., Low, T., and Hancock, N. (2021). Automatic leaf segmentation and overlapping leaf separation using stereo vision. *Array* 12, 100099. doi: 10.1016/j.array.2021.100099

Boyle, R. D., Corke, F. M., and Doonan, J. H. (2016). Automated estimation of tiller number in wheat by ribbon detection. *Mach. Vision Appl.* 27, 637–646. doi: 10.1007/s00138-015-0719-5

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "ImageNet: A large-scale hierarchical image database," in *Proceedings of Computer Vision and Pattern Recognition*. (Miami, FL, USA: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

Deng, R., Jiang, Y., Tao, M., Huang, X., Bangura, K., Liu, C., et al. (2020). Deep learning-based automatic detection of productive tillers in rice. *Comput. Electron. Agric.* 177, 105703. doi: 10.1016/j.compag.2020.105703

Dobrescu, A., Giuffrida, M. V., and Tsaftaris, S. A. (2017). Leveraging multiple datasets for deep leaf counting. In *Proceedings of the IEEE international conference on computer vision workshops*. (Venice, Italy: IEEE), 2072–2079. doi: 10.1109/ICCVW.2017.243

Fahlgren, N., Feldman, M., Gehan, M. A., Wilson, M. S., Shyu, C., Bryant, D. W., et al. (2015). A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in setaria. *Mol. Plant* 8, 1520–1535. doi: 10.1016/j.molp.2015.06.005

Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). "Multiscale vessel enhancement filtering," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention*. (Cambridge, MA, USA: Springer), 130–137. doi: 10.1007/BFb0056195

Gaggion, N., Ariel, F., Daric, V., Lambert, É., Legendre, S., Roulé, T., et al. (2021). ChronoRoot: High-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture. *GigaScience* 10, 1–15. doi: 10.1093/gigascience/giab052

Gehan, M., Fahlgren, N., Feldman, M., Wilson, M., Hill, S., Bryant, D., et al. (2015). A versatile phenotyping system and analytics platform reveals diverse temporal responses to water limitation in setaria. figshare. Dataset. doi: 10.6084/m9.figshare.1272859.v12

Gidaris, S., Singh, P., and Komodakis, N. (2018). "Unsupervised representation learning by predicting image rotations," in *Proceedings of International Conference on Learning and Representations*. (Vancouver Canada) 1–16. Available at: https://openreview.net/pdf?id=S1v4N2l0-.

Giuffrida, M. V., Minervini, M., and Tsaftaris, S. (2015). Learning to count leaves in rosette plants. In *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*. (BMVA Press) 1.1–1.13. doi: 10.5244/c.29.cvppp.1

Han, T. H., and Kuo, Y. F. (2018). Developing a system for three-dimensional quantification of root traits of rice seedlings. *Comput. Electron. Agric.* 152, 90–100. doi: 10.1016/j.compag.2018.07.001

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (Los Alamitos, CA, USA: IEEE Computer Society), 770–778. doi: 10.1109/CVPR.2016.90

Huang, Y., Cheng, Y., Bapna, A., Firat, O., Chen, D., Chen, M., et al. (2019). "Gpipe: Efficient training of giant neural networks using pipeline parallelism," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems. NY, USA*, vol. 32. Article 10, 103–112.

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2018). Squeeze-and-excitation networks. In *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, USA. 7132–7141. doi: 10.1109/CVPR.2018.00745

Kinose, R., Utsumi, Y., Iwamura, M., and Kise, K. (2022). Tiller estimation method using deep neural networks. *Reseach Square*. doi: 10.21203/rs.3.rs-1552723/v1

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Lett.* 81, 80–89. doi: 10.1016/j.patrec.2015.10.013

Miyato, T., Maeda, S.-I., Koyama, M., and Ishii, S. (2019). "Virtual adversarial training: A regularization method for supervised and semi-supervised learning," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (IEEE), Vol. 41, 1979–1993. doi: 10.1109/TPAMI.2018.2858821

Noroozi, M., and Favaro, P. (2016). "Unsupervised learning of visual representations by solving jigsaw puzzles," in *Proceedings of European Conference on Computer Vision*. (Amsterdam, The Netherlands: Springer), 69–84. doi: 10.1007/978-3-319-46466-4_5

Noroozi, M., Pirsiavash, H., and Favaro, P. (2017). "Representation learning by learning to count," in *Proceedings of Internatinal Conference on Computer Vision*. (Venice, Italy: IEEE), 5898–5906. doi: 10.1109/ICCV.2017.628

Noroozi, M., Vinjimoor, A., Favaro, P., and Pirsiavash, H. (2018). "Boosting self-supervised learning *via* knowledge transfer," in *Proceedings of Computer Vision and Pattern Recognition*. (alt Lake City, UT, USA: Springer), 9359–9367. doi: 10.1109/CVPR.2018.00975

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., and French, A. P. (2017). "Deep learning for multi-task plant phenotyping," in *Proceedings of 2017 IEEE International Conference on Computer Vision Workshops*. (Venice, Italy: IEEE), 2055–2063. doi: 10.1109/ICCVW.2017.241

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). "Facenet: A unified embedding for face recognition and clustering," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Boston, MA, USA: IEEE), 815–823. doi: 10.1109/CVPR.2015.7298682

Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in *Proceedings of International Conference on Learning Rresentations*. (San Diego, CA, USA). Available at: https://arxiv.org/abs/1409.1556.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). "DeepFace: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (Columbus, OH, USA: IEEE), 1701–1708. doi: 10.1109/CVPR.2014.220

Ubbens, J., Cieslak, M., Prusinkiewicz, P., and Stavness, I. (2018). The use of plant models in deep learning: An application to leaf counting in rosette plants. *Plant Methods* 14, 1–10. doi: 10.1186/s13007-018-0273-z

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: A deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01190

Utsumi, Y., Nakamura, K., Iwamura, M., and Kise, K. (2019). "DNN-based tiller number estimation for coping with shortage of labeled data," in *Proceedins of Computer Vision Problems in Plant Phenotyping*, Vol. 1. (Long Beach, CA, USA). Available at: https://www.plant-phenotyping.org/lw_resource/datapool/systemfiles/elements/files/42e1416f-70a4-11e9-b1c5-dead53a91d31/current/document/UtsumiCVPPP2019.pdf.

Wang, T., Rostamza, M., Song, Z., Wang, L., McNickle, G., Iyer-Pascuzzi, A. S., et al. (2019). SegRoot: A high throughput segmentation method for root image analysis. *Comput. Electron. Agric.* 162, 845–854. doi: 10.1016/j.compag.2019.05.017

Ward, D., Moghadam, P., and Hudson, N. (2018). "Deep leaf segmentation using synthetic data," in *In Proceedings of the British Machine Vision Conference (BMVC) Workshop on Computer Vision Problems in Plant Pheonotyping (CVPPP)*. (Newcastle, UK). Available at: http://bmvc2018.org/contents/workshops/cvppp2018/0026.pdf.

Wu, D., Wu, D., Feng, H., Duan, L., Dai, G., Liu, X., et al. (2021). A deep learning-integrated micro-CT image analysis pipeline for quantifying rice lodging resistance-related traits. *Plant Commun.* 2. doi: 10.1016/j.xplc.2021.100165

Wu, J., Wu, Q., Pagès, L., Yuan, Y., Zhang, X., Du, M., et al. (2018). RhizoChamber-monitor: A robotic platform and software enabling characterization of root growth. *Plant Methods* 14, 1–15. doi: 10.1186/s13007-018-0316-5

Xie, Q., Mayes, S., and Sparkes, D. L. (2015). Optimizing tiller production and survival for grain yield improvement in a bread wheat × spelt mapping population. *Ann. Bot.* 117, 51–66. doi: 10.1093/aob/mcv147

Yasrab, R., Atkinson, J. A., Wells, D. M., French, A. P., Pridmore, T. P., and Pound, M. P. (2019). RootNav 2.0: Deep learning for automatic navigation of complex plant root architectures. *GigaScience* 8. doi: 10.1093/gigascience/giz123

Zhang, R., Isola, P., and Efros, A. A. (2016). "Colorful image colorization," in *Proceedings of European Conference on Computer Vision*. (Amsterdam, The Netherlands: Springer), 649–666. doi: 10.1007/978-3-319-46487-9_40

frontiers | Frontiers in Plant Science

TYPE Systematic Review
PUBLISHED 26 January 2023
DOI 10.3389/fpls.2022.1084847

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA), United States

REVIEWED BY
Naveen Kumar Mahanti,
Dr.Y.S.R. Horticultural University, India
Ebrahiema Arendse,
Agricultural Research Council of South Africa (ARC-SA), South Africa

*CORRESPONDENCE
Zhenhui Ren
renzh68@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 31 October 2022
ACCEPTED 21 December 2022
PUBLISHED 26 January 2023

CITATION
Liu X, Li N, Huang Y, Lin X and Ren Z (2023) A comprehensive review on acquisition of phenotypic information of Prunoideae fruits: Image technology.
Front. Plant Sci. 13:1084847.
doi: 10.3389/fpls.2022.1084847

COPYRIGHT
© 2023 Liu, Li, Huang, Lin and Ren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A comprehensive review on acquisition of phenotypic information of Prunoideae fruits: Image technology

Xuan Liu[1†], Na Li[1†], Yirui Huang[2], Xiujun Lin[1] and Zhenhui Ren[1*]

[1]College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, China, [2]College of Information Engineering, Hebei GEO University, Shijiazhuang, China

Fruit phenotypic information reflects all the physical, physiological, biochemical characteristics and traits of fruit. Accurate access to phenotypic information is very necessary and meaningful for post-harvest storage, sales and deep processing. The methods of obtaining phenotypic information include traditional manual measurement and damage detection, which are inefficient and destructive. In the field of fruit phenotype research, image technology is increasingly mature, which greatly improves the efficiency of fruit phenotype information acquisition. This review paper mainly reviews the research on phenotypic information of Prunoideae fruit based on three imaging techniques (RGB imaging, hyperspectral imaging, multispectral imaging). Firstly, the classification was carried out according to the image type. On this basis, the review and summary of previous studies were completed from the perspectives of fruit maturity detection, fruit quality classification and fruit disease damage identification. Analysis of the advantages and disadvantages of various types of images in the study, and try to give the next research direction for improvement.

KEYWORDS

Prunoideae fruits, spectral image, phenotypic information, nondestructive techniques, image technology

1 Introduction

Fruit phenotype describes the expression of fruit traits. Research on fruit traits can be done at multiple levels, including cells, tissues, organs, individual fruits, whole plant fruits, and even entire orchards (Dhondt et al., 2013). Phenotypic information of fruit includes but is not limited to geometric size, biomass content, moisture content, and skin color (Shakoor et al., 2017; Choudhury et al., 2019). The variation of phenotypic

information is closely related to the market value of fruits. Therefore, it is of great significance to obtain accurate fruit phenotype information for maximizing the economic value of fruits (Mahlein, 2016). Fruit harvesting is an important part of agricultural production. The completion of harvesting at the appropriate fruit harvesting window is the basis for ensuring consumers to obtain high quality fruits. By obtaining phenotypic information such as fruit hardness and SSC (Sohaib Ali Shah et al., 2020), the fruit harvesting window can be accurately grasped, thus guiding the fruit harvesting work. In addition, the research on fruit quality evaluation (Qin et al., 2013; Su and Sun, 2018), fruit disease and damage (Ali et al., 2019) based on phenotypic information is also very meaningful.

In traditional methods, digital calipers and electronic scales were used to measure the size and weight of fruits (Zhang et al., 2014), and Folin-Ciocalteu method (total polyphenol content) and hand-held refractometer (sugar content) (Pissard et al., 2013; Kopjar et al., 2017) were used to determine the polyphenol and sugar content of fruits respectively. These methods are meaningful, but the disadvantages are also obvious, such as the measurement process is time-consuming and damages the integrity of the fruit. With the rapid technological advancement in electronics and computers sectors different technologies were developed to obtain fruit phenotypic information efficiently, accurately and non-destructively. Compared with the traditional detection technology, the phenotypic information detection research based on spectral technology realizes the non-destructive detection of fruit according to the difference between the light absorption rate and the reflectivity inside the fruit (Toivonen et al., 2017). Due to the rich spectral data and image information contained in the images (Fernández-Novales et al., 2019), multispectral and hyperspectral imaging devices have significantly improved the efficiency of fruit phenotype acquisition, and have been widely used by researchers in the related research of fruit phenotype information. The wide applicability of RGB imaging equipment makes the study of phenotypic information based on RGB images a new research hotspot (Blasco et al., 2017). In addition, thermal imaging technology was used in the study of fruit temperature (Osroosh and Peters, 2019; Ranjan et al., 2022), and computer tomography technology (Kritzinger et al., 2017; Karmoker et al., 2018) and laser backscatter imaging technology (Adebayo et al., 2016; Mozaffari et al., 2022) were used in the study of fruit internal quality detection. The phenotypic information acquisition technology based on image technology avoids the measurement error caused by subjective factors in traditional detection methods (Fu et al., 2020), and further improves the accuracy of phenotypic information acquisition. As shown in Table 1, the advantages and disadvantages of imaging techniques in the acquisition of fruit phenotypic information and related research were summarized.

The expression of phenotypic information in fruits may change at different stages of ripening and different kinds of diseases. The researchers found that fungal infection caused an abnormal increase in water content due to the breakdown of carbohydrates (Sun et al., 2018) and a subsequent decrease in chlorophyll content during fruit ripening (Muhua et al., 2007; Lleó et al., 2009). The changes of fruits are closely related to the phenotypic information they present. Therefore, researchers have completed the relevant research on fruits while obtaining phenotypic information based on different types of images.

This paper reviews the phenotypic information acquisition and related research based on RGB images, hyperspectral images, and multispectral images, of Prunoideae fruits. The chapter classification is completed according to the image type, and the relevant literature is reviewed and summarized according to the research purposes of fruit quality classification (Pu et al., 2015), disease damage identification (Shao et al., 2019), and maturity detection (Sohaib Ali Shah et al., 2020). The advantages and disadvantages of the completed studies were discussed. Finally, the future trends and challenges of phenotypic information acquisition based on image technology were prospected. Based on the high-frequency words of the cited papers, the word frequency distribution map of the references is drawn. It can be seen from the color degree of the key words in the figure that the related research based on hyperspectral

TABLE 1 Summary table of imaging technology in the acquisition of fruit phenotypic information and related research.

Research object	Imaging technology	Imaging environment	Portability	Ref.
Cherry	Commercial thermal-RGB sensor	Outdoor environment	Portability	(Ranjan et al., 2022)
Plum	Laser backscattering imager	Closed darkroom t	Non-portability	(Rezaei Kalaj et al., 2016)
Apricot	Laser backscattering imager	Closed darkroom	Non-portability	(Mozaffari et al., 2022)
Peach	Multispectral image system	Closed darkroom	Non-portability	(Herrero-Langreo et al., 2011)
Cherry	UAV Multispectral Imager	Outdoor environment	Portability	(Karydas et al., 2020)
Plum	Hyperspectral imaging system	Laboratory environment	Non-portability	(Li et al., 2018a)
Nectarine	Hyperspectral imaging system	Closed darkroom	Non-portability	(Munera et al., 2018)
Plum	Mobile phones, cameras	Outdoor environment	Portability	(Ahmad et al., 2020)
Cherry	Digital camera	Laboratory environment	Non-portability	(Momeny et al., 2020)

images and RGN images accounts for the majority in quantity. As shown in Figure 1, the keyword frequency distribution map of the references in this paper.

2 Phenotypic information acquisition and related applications based on RGB image

RGB mode is a color standard, by changing the red (R), green (G), blue (B) three color channels and their superposition to get a variety of colors. RGB mode is one of the most widely used color systems, and RGB images can provide data information such as color features, texture features, and geometric shapes of fruits (Kaur et al., 2018). Compared with multispectral images and hyperspectral images, RGB images can be acquired by smartphones, cameras and other means, and the acquisition methods are more diversified and universal. At the same time, the image acquisition equipment has low requirements on the acquisition environment, even in cloudy days, sunshine or indoor environment (Miragaia et al., 2021).

In recent years, researchers have completed many related studies based on fruit RGB images. This section reviews the research progress of fruit maturity detection, disease damage identification and fruit quality classification based on RGB images.

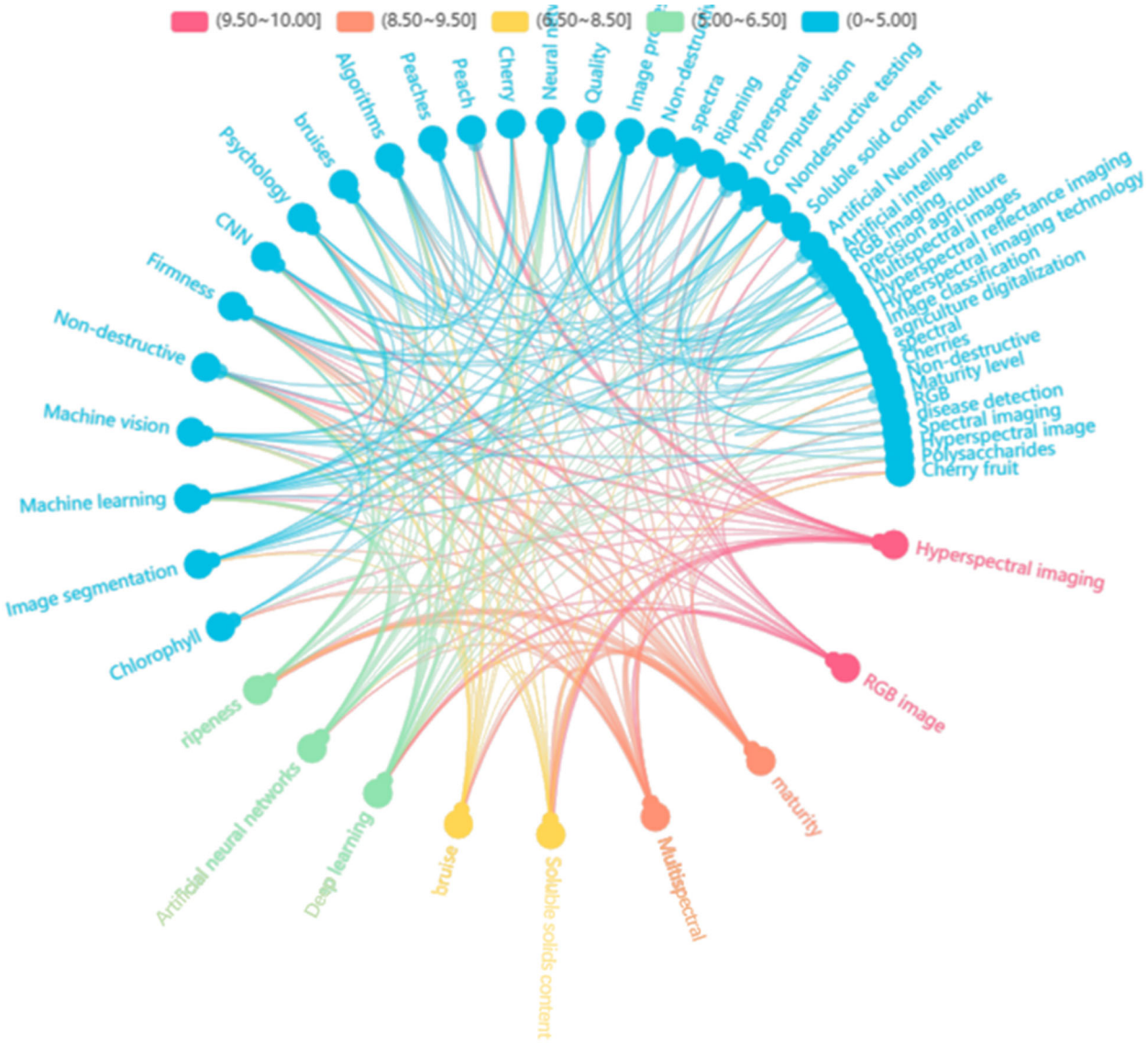


FIGURE 1
Reference keyword frequency distribution map.

2.1 Fruit maturity detection based on RGB image

As the fruit gradually mature chlorophyll degradation, anthocyanins or carotenoids and other new pigments began to synthesize, resulting in fruit color began to change. Therefore, consumers usually associate color with the ripening stage of fruits (Blasco et al., 2017).

In 2015, S. Taghadomi et al. completed the determination of color parameters during cherry ripening (Taghadomi-Saberi et al., 2015). A CCD camera (PROLINE UK, Model 565s with 510 by 492 pixels resolutions, London, United Kingdom) was used to acquire cherry images. Threshold segmentation technology and Otus algorithm are used to extract cherry image. After screening 37 common features using PCA, 7 features were obtained as input vectors. The relationship between the L* a* b* value measured by the colorimeter and the color features extracted from the cherry image was modeled and analyzed based on the artificial neural network (ANN) using MATLAB. The Levenberg-Marquardt algorithm and trainbr function were used to train the network. The results showed that the ANN with structure of 7-14-11-3 had the best modeling effect on L* a* b* color parameters during cherry ripening (R^2 = 0.9999). In 2018, Indian scholar Kaur et al. completed a study on the evaluation of fruit maturity based on RGB images of plums (Kaur et al., 2018). The plum images were captured using a digital camera (Nikon Coolpix S3200, Resolution-4608 × 3456, and Format-JPEG) under natural light. In their research, the uniform threshold operator in MATLAB image processing toolbox is used to complete image segmentation.The average RGB value is used to extract color features, and entropy, local binary pattern and discrete cosine transform are used to extract texture features. Based on the multi-attribute decision making (MADM) theory, the decision of maturity level is completed. The results show that the developed system accurately determines the maturity level of plums. The correlation strength between color features and texture features and maturity at different stages is shown in the article (Kaur et al., 2018).

In the following year, researcher Mostafa Khojastehnazhand et al. completed maturity detection and volume estimation based on RGB images of apricot fruit (Khojastehnazhand et al., 2019). They obtained seven kinds of feature information of apricot images, and finally selected G channel, gray value, L* and b* as input features. Discriminant analysis models based on LDA and Quadratic Discriminant Analysis (QDA) are compared. The results show that R^2 values of LDA and QDA are 0.904 and 0.923 respectively.

There are many ways to obtain RGB images, and the requirements for the environment are not high. However, different acquisition environment and acquisition equipment will make the image have some differences, such as the impact of light conditions, image quality differences. Miragaia et al. completed the research of deep learning algorithm based on convolutional neural network (CNN) to analyze plum maturity detection in real environment (Miragaia et al., 2021). The method of obtaining plum fruit images used in this study is not limited. Plum fruit images obtained by devices such as smartphones and cameras can be used for maturity recognition. The influence of different image quality is solved, so that the obtained images can be used for the detection of fruit maturity. The results show that the developed system has good robustness. Even if the image has different illumination conditions and focus, it can correctly complete the classification of plum fruit maturity, and the efficiency is above 94%. Table 2 summarizes the studies on fruit maturity detection based on phenotypic information using RGB images.

It can be concluded from Table 2 that the related research on maturity detection and classification completed by researchers using RGB images has achieved good research results. The reason for the analysis is that the plum fruits studied are all varieties with strong correlation between maturity and fruit color changes, and RGB images can better reflect the color characteristics of fruits at different maturity stages.

2.2 Fruit quality detection based on RGB image

In the field of fruit quality detection, the quality detection based on RGB images is more focused on the phenotypic information obtained from the fruit surface. This is because the fruit RGB images only contain surface phenotypic information such as color features, geometric features, and texture features. We try to give the steps followed by common image processing in target detection based on RGB images (Figure 2). It should be pointed out that the processing steps mentioned are not necessarily used, and the specific methods are related to the actual needs.

Iranian researcher Esehaghbeygi et al. used a high-resolution CCD (PROLINE UK, Model 565s with 510 by 492 pixels resolution) camera to obtain RGB images of peach fruit, and completed the study of color grading and size evaluation of peach fruit (Esehaghbeygi, 2010). An edge detection algorithm was developed to estimate the volume of peach fruit, and the Mesh function in MATLAB software was used to obtain the hue, saturation and value of peach image and complete the classification of peach fruit color and the detection of surface spots. The results show that the detection rate of white spots is 96.7%, the detection rate of brown spots is 85%, the accuracy of size classification is 96%, and the accuracy of color classification is 90%. It is an unavoidable problem to eliminate the influence of other light in the detection of fruit quality in outdoor environment. Wang et al. completed a study on the color

TABLE 2 Review on fruit maturity detection based on phenotypic information using RGB images.

Research object	Preprocessing	Method	Imaging device	Result	Ref.
Cherry	Thresholding and Otsu's algorithm techniques	ANN	CCD camera	R^2 = 0.9999	(Taghadomi-Saberi et al., 2015)
Plum	Uniform thresholding operator	MADM	Digital camera	Correlation between Acidity and Green: R^2 = 0.9966; Correlation between R/G and SSC: R^2 = 0.8464	(Kaur et al., 2018)
Apricot	Averaging filter	LDA and QDA	Digital camera	LDA: R^2 = 0.904. QDA: R^2 = 0.923	(Khojastehnazhand et al., 2019)
Plum	Mean image subtraction, mean pixel subtraction	Deep Learning	Mobile phone, camera	Maturity recognition accuracy>94%	(Miragaia et al., 2021)

evaluation of sweet cherry (Wang et al., 2012). The acquisition of cherry images was completed by a digital camera (Nikon D5000 Nikon Inc., Melville, NY, USA). The acquisition environment was an outdoor environment with sufficient light. The cherry images of direct sunlight, bright shadow and dark shadow were obtained respectively. The distance between the camera and the cherry was 0.5 m. Using image processing technology to eliminate the strong light spots caused by light, mainly includes two steps: (1) using the green channel of color grading area to detect the image pixels with dazzling

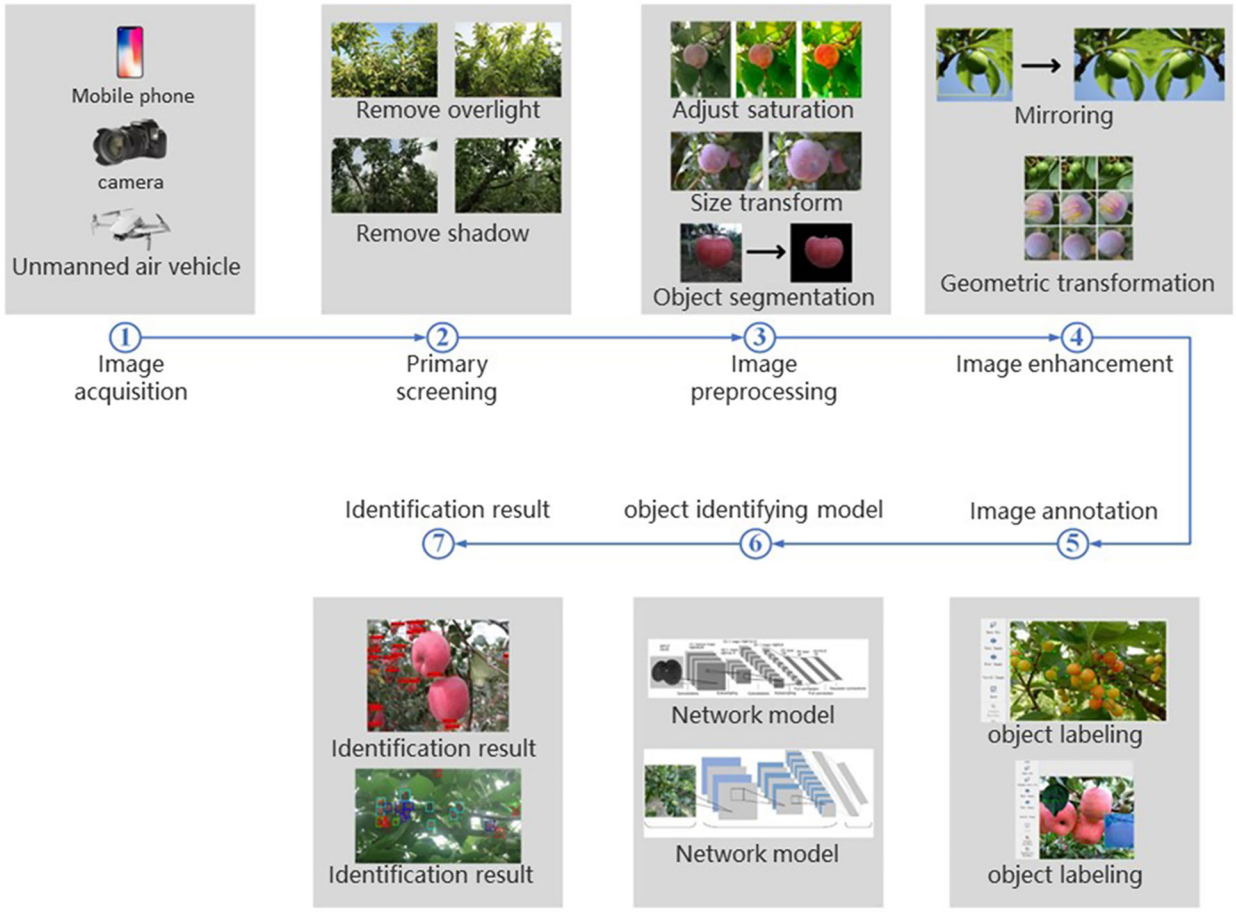


FIGURE 2
Image processing steps in target recognition.

reflection; (2) Eliminate detected pixels from the color rating area of the image. In the study, a color rating system was successfully developed to complete the color rating of cherry images. The results showed that the overall accuracy of color rating was more than 85%.

Taghadomi-Saberi 's team combined image processing technology and ANN to complete the evaluation of antioxidant activity and anthocyanin content of sweet cherries (Taghadomi-Saberi et al., 2014). A color CCD camera (PROLINE UK, Model 565S, London, United Kingdom) was used to obtain the cherry image. The Otsu method is used to complete the cherry image segmentation. Two prediction models based on ANN and adaptive neuro-fuzzy inference system (ANFIS) were established and compared, and the accuracy of the two models for antioxidant activity and anthocyanin content was evaluated. The results showed that the prediction models with the structure of 11-14-9-1 and 11-6-20-1 based on ANN had the highest correlation coefficients with antioxidant activity and anthocyanin content, which were R = 0.93 and R = 0.98, respectively. Meanwhile, the ANFIS prediction model obtained the best prediction results (R = 0.87 and R = 0.90) when using triangular and two-term Gaussian membership functions. Cherry fruit skin color as an important indicator to measure its maturity and quality has become a research hotspot. Researchers (Wang et al., 2012; Taghadomi-Saberi et al., 2014; Taghadomi-Saberi et al., 2015) have studied color as an important indicator in their studies, but have neglected the effect of cherry shape appearance on its quality. In 2020, Mohammad Momeny 's team completed the analysis of cherry quality from the perspective of cherry shape appearance (Momeny et al., 2020). They used an improved CNN to detect the appearance shape of cherries. The classification algorithm completed the classification of regular and irregular cherries with an accuracy of 99.4%. By using mirroring, rotation and other methods to expand the data set, 14,380 images are finally obtained. It is necessary to establish a perfect data set to obtain higher classification accuracy. In the study, the deep CNN based on hybrid pooling was compared with KNN, ANN, Fuzzy, and integrated decision tree methods based on histogram of gradient and local binary pattern feature extraction methods. The results show that the improved CNN method is suitable for the detection of cherry appearance (regular and irregular shaped).

In the review of reported studies, we found that building a complete and sufficient image data set is a prerequisite for obtaining accurate prediction results. Villacrés, a researcher in Chile, installed an RGB camera on a tractor through a self-stabilized pan-tilt for image acquisition. The camera is 1m away from the cherry tree, and a total of 15,000 images are obtained for the construction of the data set, which greatly improves the efficiency of image acquisition (Villacrés and Cheein, 2020). Four different color borders were used to label the cherries in advance and different colors were used to distinguish the size of the cherries. Faster R-CNN meta-architecture and Inception V2 were used as feature extractors for the detection of cherries. The results show that the cherries in the image can be identified with an accuracy of 85% and the cherries are divided into four sizes according to the requirements of farmers. The main reason why cherries are not detected in the study is that the fruit pixels caused by occlusion are too small to detect (pixels less than 20). Table 3 summarizes the studies on fruit quality detection based on phenotypic information using RGB images.

Through the analysis of Table 3, we can conclude that the current research on fruit color quality grading and fruit size estimation based on the color features and geometric size features of RGB images has achieved high accuracy. In the above research, the recognition and classification accuracy based on CNN is more satisfactory. Researchers have further improved the research accuracy based on CNN by innovating on the infrastructure. However, in the research of small fruit recognition with small fruit volume and complex image acquisition environment, improving the recognition detection rate is still the focus of the next research.

TABLE 3 Research summary of fruit quality detection based on phenotypic information using RGB images.

Fruit types	Imaging environment	Research aim	Research method	Result	Ref.
Peach	Standard Lighting Room	Color rating; Volume estimates	Edge detection algorithm	Detection rate of white and brown spots: 96.7%, 85%; Size and Color Classification Accuracy: 96%, 90%	(Esehaghbeygi, 2010)
Cherry	Outdoor with sufficient light	Color rating;	Color Rating System	Color Rating Accuracy: >85%	(Wang et al., 2012)
Cherry	Uniform illumination environment	Antioxidant activity; Anthocyanin content	ANN, ANFIS	Triangular: R=0.87; Gaussian membership functions: R=0.90	(Taghadomi-Saberi et al., 2014)
Cherry	Inside the lighting box	Appearance shape classification	CNN	Classification accuracy: 99.4%	(Momeny, et al., 2020)
Cherry	Outdoor in different light conditions	Identification and size grading	Faster R-CNN	Recognition rate: 85%	(Villacrés and Cheein, 2020)

2.3 Fruit disease damage detection based on RGB image

CNN method widely used in research related to visual recognition, such as image classification (Ahmad et al., 2017), target detection and recognition (Ahmad et al., 2018b), and image matching (Ahmad et al., 2018a). In the previous chapters, some researchers have also reported the use of CNN to complete related research on fruit quality. This chapter summarizes the related research on fruit disease image classification and recognition based on RGB image, including fruit disease image acquisition method, image preprocessing technology and neural network algorithm. The feature information contained in RGB images is mostly the phenotypic information existing in the fruit surface, such as fruit color, size and other feature data. Therefore, the fruit disease detection based on RGB images is mainly aimed at the disease and damage of the fruit surface layer.

In a study on plum disease detection reported in 2020, researcher Ahmad et al. proposed a plum disease detection framework based on CNN (Ahmad et al., 2020). Four structures are compared in the study: AlexNet, VGG-16, Inception and Resnet. The data set is expanded and made more challenging through data augmentation to achieve robust model training. The recognition accuracy of the model before and after data enhancement is compared. The experimental results show that the performance of the model increases with the increase of the number and complexity of the data set. The research on different architectures shows that the disease identification and classification results based on Inception-V3 model are the best, reaching 92%. In order to run on resource-constrained devices, Jamil Ahmad et al. quantized the Inception-v3 model from FP32 precision to FP16, gaining a 2× speedup and 2× less memory requirement.

Huang et al. completed a peach fruit RGB image disease detection method based on asymptotic non-local means (ANLM) image algorithm and parallel convolutional neural network fusion in 2020 (Huang et al., 2020). Firstly, ANLM is used to remove the interference of complex background in the image, and then parallel convolutional neural network is used to identify peach disease features. In the research, the improved elu activation function is used to replace the traditional Eelu activation function, and the linear particle swarm optimization ELM proposed in the research is used to replace the traditional softmax layer. Through the improvement of the algorithm, the convergence speed and accuracy of the network are significantly improved. The results showed that the accurate detection rates of brown rot, black spot, anthracnose, scab and normal peach were all above 85%, indicating that the improved parallel convolutional neural network algorithm was an effective method for peach disease detection.

The segmentation algorithm greatly affects the segmentation accuracy, and accurately obtaining the region of interest is the premise of accurately identifying the type of disease defects. Therefore, Alosaimi et al. proposed a new CNN model for detecting peach disease categories (Alosaimi et al., 2021). In the study, Mask R-CNN was used to complete the segmentation of the diseased area, and then the VGG-19 architecture was used to identify the type of the segmented area. The peach disease database in the study consists of a public database and photos obtained in the natural environment. The photos obtained in the natural environment include different types such as direct sunlight and cloudy days. Finally, the mean Average Precision (mAP) was used to evaluate the performance of the model. The results showed that the improved CNN for peach disease classification had mAP = 94%.

Deep learning is widely used in related research using imaging data to detect disease categories, but it is difficult to collect a large number of peach disease images, and the sample images are unbalanced. In response to this problem, Yao et al. proposed an improved Xception network called L2MXception (Yao et al., 2021). The network integrates regularization terms of L2 norm and mean. In the study, the recognition results of seven deep learning models were compared, and the composition of the peach disease image dataset included seven disease types. The results show that the classification accuracy of L2MXception reaches 93.85%, which is 28.48% higher than that of Xception model.

Based on the above review, we found that various machine learning algorithms have been applied to the research field of fruit tree disease recognition. Compared with the traditional technology, the application of new technology greatly improves the detection speed and recognition classification accuracy. Obtaining rich and complete data sets is necessary for the establishment of models with excellent performance. This puts forward higher requirements for the acquisition efficiency of image data sets, and automatic, efficient and reliable data acquisition equipment becomes very important.

2.4 Other related research based on fruit RGB image

Based on the feature information obtained from RGB images, researchers have completed related research on fruit maturity, quality detection, disease damage and other fields. In addition, it was also reported that relevant researchers completed other related studies based on fruit RGB images.

In 2021, Ropelewska completed the study of cherry variety discrimination using the acquired cherry images (Ropelewska et al., 2021). The research proves that the discriminant model based on texture parameters obtained from different color channels and texture parameters obtained from different color spaces has high recognition accuracy for cherry varieties. Discriminant models based on histogram, co-occurrence

matrix, run-length matrix, autoregressive model and gradient map are considered in the experiment. The results show that the accuracy of the model based on the texture parameters selected from the color space is slightly higher than that of the texture parameters obtained by inputting the color channel. In the color channel R, X and color space, the accuracy of the texture parameters based on histogram and co-occurrence matrix to distinguish the three sweet cherries reached 100%. In the color channel, the histogram model based on the color channel L produces the highest accuracy of 97%. Similar studies have also been applied to the field of peach variety identification. Ropelewska et al. compared the classification results of different color channels and different discriminant models by acquiring images of peach skin, flesh, stone and seed texture. (Ropelewska and Rutkowski, 2021). The results show that the texture features based on different color channels can better complete the identification of peach varieties.

3 Phenotypic information acquisition and related applications based on hyperspectral image

Hyperspectral imaging technology has broad application space in the field of agriculture, and has gradually become one of the important and cutting-edge technical means in agricultural applications. The application of hyperspectral images includes crop growth monitoring, crop stress monitoring (Chattaraj et al., 2013), crop yield primary estimation (Ye et al., 2006), vegetation coverage monitoring, and non-destructive monitoring (Huang et al., 2014) of agricultural products, providing service support for precision agriculture and agricultural management. Compared with multispectral images, hyperspectral imaging equipment has a wider range of imaging wavelengths, which greatly improves the information richness of hyperspectral images. In the processing technology and application, the acquisition of rich spectral data makes more reasonable and more effective analysis possible. Therefore, hyperspectral image technology has incomparable development potential. Hyperspectral imaging equipment is mainly composed of halogen light source, imaging lens, computer, transmission platform, transmission motor and other parts.

3.1 Fruit maturity and biochemical parameters detection based on hyperspectral image

Hyperspectral imaging equipment obtains spectral data of hundreds or thousands of samples with nanoscale sampling resolution. There is inevitably a large amount of redundant data in the rich spectral data, and more redundant data usually lead to longer and unreliable prediction of dependent variables. Therefore, eliminating too much redundant data has become an inevitable process. Researchers have used many variable selection methods, such as ANN, partial least squares (PLS), principal component analysis (PCA), genetic algorithms (GA), etc. For details on variable selection methods, please refer to Xiaobo et al. (Xiaobo et al., 2010). There is no accurate range for the research of fruit quality, and the existing related research focuses on fruit SSC, acidity, firmness, etc. These parameters are directly related to the fruit flavor tasted by consumers.

Munera et al. obtained hyperspectral images of peach fruit ripening process with a wavelength range of 450-1040 nm using laboratory hyperspectral imaging equipment (Munera et al., 2017). Internal Quality Index (IQI) and Ripening Index (RPI) were introduced to evaluate peach fruit maturity. Variable Importance in Projection (VIP) was used to complete the spectral screening of hyperspectral images, and the regression models of IQI and RPI were established based on PLS regression analysis. Based on the spectral data of each pixel, the IQI and RPI prediction values of the peach fruit image were obtained, and the visualization of the peach fruit maturity distribution map was realized. The results showed that in the prediction of IQI and RPI of the two varieties of peach fruit, the R^2 values of the two indexes and the two varieties were greater than 0.87. Li et al. completed a study on the detection and classification of SSC and PH content and maturity of cherry fruits based on near-infrared hyperspectral imaging technology (Li et al., 2018e). The wavelength range of hyperspectral image of cherry fruit is 874-1734 nm, and the distance between lens and sample is 30.5 cm. In order to accurately complete the classification of cherry maturity, the samples were classified by five orchard owners according to the principle of majority, and the classification model of cherry maturity was established by supervised learning method Linear Discriminant Analysis (LDA). According to the spectral characteristic parameters of different maturity categories, the classification of cherry maturity was completed. The results show that the accuracy of the classification model is 96.4%.

The relevant literature shows that the research methods based on hyperspectral images have been widely used in the research of fruit maturity, including the optical index (Lleó et al., 2011) based on spectral image acquisition for maturity detection and grading, and the maturity prediction and grading based on the established prediction model (Munera et al., 2017; Li et al., 2018e). The summary of related research in recent years is completed through Table 4.

From the Table 4, we found that the wavelength range used in the study was mostly concentrated near the chlorophyll absorption peak and SSC absorption peak. This is because the content of biochemical substances in the fruit will change with the change of maturity. The feasibility of fruit maturity classification by detecting the content of biochemical

substances has been verified by relevant researchers. However, there are some differences in wavelength selection among different varieties.

In 2017, Zhu et al. obtained hyperspectral images of different slices inside peaches using visible and short-wave near-infrared spectral imaging equipment (380-1030 nm) and long-wave near-infrared spectral imaging equipment (874-1734 nm) (Zhu et al., 2017). Savitzky-Golay smoothing and standard normal variate transformation were used to preprocess the spectral images. Subsequently, PLSR and LS-SVM modeling methods were established and compared. In their study, the contents of protopectin, water-soluble pectin and total pectin were predicted respectively, and results showed that the prediction model had better prediction results for protopectin than for water-soluble pectin. According to this result, it was analyzed that water-soluble pectin dissolved in the process of destructively obtaining peach slices. The visualization of peach fruit pectin content distribution was also completed in the study. Pectin content is considered to be closely related to maturity. Pectin distribution map can be used as an indicator to provide guidance for understanding fruit ripening and post-harvest storage systems. In the subsequent study, Li et al. completed the detection of non-destructive quality attributes of plum fruit from three quality indicators: color, firmness and SSC (Li et al., 2018a). Two hyperspectral cameras with wavelength ranges of 600-975 nm in visible and near infrared (VNIR) region and 865-1610 nm in short wave near infrared (SWIR) region were used to obtain hyperspectral images of two varieties of plum fruits. Because the surface of plum fruit is smooth and has strong reflection, which is different from peach fruit, the intensity of fruit edge in the sample spectral image is low. They cited the automatic correction method for light scattering of spherical objects developed by Gomez-Sanchis et al. to improve this phenomenon (Gómez-Sanchis et al., 2008). At the same time, a PLSR model was established for the non-destructive measurement of firmness, SSC and color components of two different plum varieties. The results showed that there was a strong correlation between SSC and SWIR spectra, and the predicted correlation coefficient r_p^2 was greater than 0.8. The VNIR spectrum has a good correlation with color, and the r_p^2 value is greater than 0.7 for L* and a*. Both hyperspectral imaging systems have low prediction accuracy for hardness.Shen et al. predicted the SSC of green plum using sparse autoencoder (SAE) in a study reported in 2020 (Shen et al., 2020). SAE is an unsupervised machine learning algorithm, which continuously adjusts the autoencoder parameters and finally completes the model training by calculating the error between the autoencoder output and the original input. The hyperspectral imaging equipment used in the experiment has a spectral range of 400-1000 nm and a spectral resolution of 2.8 nm. A multi-layer network model SAE-PLSR was proposed to predict SSC, and the sparsity parameter ρ was set to 0.01 (Tang et al., 2016). At the same time, BP, SVR, PLSR, SAE-BP, SAE-SVP and SAE-PLSR were compared. The results showed that SAE-PLSR model had the best prediction results, r and root mean square error (RMSE) of prediction set were 0.938 and 0.654 respectively.

With the improvement of deep learning theory, Yang et al. introduced deep learning theory into the prediction of SSC content of peach, and proposed a SSC estimation method of fresh peach based on deep features of hyperspectral image fusion information (Yang et al., 2020). The distance from the peach sample to the lens is 220 mm, the wavelength range is 900-1740 nm, and the spectral resolution is 5 nm. In the research, the stacked autoencoder is used for the depth feature of the spectrum and the depth feature extraction of the image. Unlike the sparse autoencoder used by Shen et al. (Shen et al., 2020), the stacked autoencoder is a cascade of multiple autoencoders to complete the task of layer-by-layer feature extraction. The resulting features are more representative and have a small dimension. Finally, a stack autoencoder - random forest peach SSC estimation model based on hyperspectral image fusion information depth features was established. The results show that the estimation model with the network structure of 1237-650-310-130 has the highest accuracy, the training set R^2 = 0.9184, and the validation set R^2 = 0.8838. In 2022, Xuan et al. completed the analysis of SSC, firmness, diameter, weight and other internal and external quality of peach fruit based on hyperspectral images (Xuan et al., 2022). The wavelength range of hyperspectral image is 400-1000 nm, and the distance between the sample and the lens is set to 47 cm. The CARS and random frog algorithms were used to select effective wavelengths, and a non-destructive regression model for

TABLE 4 Summary of research on maturity detection based on hyperspectral imaging.

Research Objective	Research object	Method	Wavelength coverage	Wavelength screening	Optical index	Model	Ref.
Fruit Maturity Detection of Prunoideae	Peach	Optical index	640, 675, 680, 720, 730, 800 nm	/	Ind1, Ind2, Ind3, IAD	/	(Lleó et al., 2011)
	Peach	Forecasting model	380-1030 nm	SPA, UVE, CARS	/	PLSR, LS-SVM	(Zhu et al., 2016)
	Peach		450-1040 nm	VIP	/	PLS	(Munera et al., 2017)
	Cherry		874-1734 nm	GA, SPA	/	LDA	(Li et al., 2018e)

predicting SSC and firmness was established based on multiple linear regression (MLR). The results showed that the CARS-MLR model had a good prediction effect on SSC, and the training set and validation set were R^2_C 0.856 and R^2_V 0.841, respectively. In the study, the estimation information of fruit size and weight was obtained by extracting pixel diameter and area, the estimation of peach diameter was completed by using the minimum boundary rectangle method, and the weight of peach fruit was predicted by MLR model. The results showed that the maximum error was 3.14 mm, the average absolute error was 0.94 mm and the average percentage error was 1.01%. The MLR regression model was established for weight estimation. The training set and validation set were $R^2_C = 0.946$ and $R^2_V = 0.957$, respectively.

Li et al. completed a study on the detection and classification of SSC and PH content and maturity of cherry fruit based on near-infrared hyperspectral imaging technology (Li et al., 2018e). The research content related to maturity detection and classification has been summarized in the previous chapters. Here, the related research on SSC and PH content is reviewed. The cherry samples studied completed the establishment of calibration set and validation set according to the principle of interval sampling. Principal components regression model and PLSR model based on full spectrum and GA-MLR model based on characteristic bands were established. For the prediction model established by full spectral band, the prediction result of SSC is better than that of PH, the reason is that the fluctuation range of PH value is too small, and the PH value only fluctuates between 3.3 and 4.1. The GA and SPA wavelength selection algorithm are compared in establishing the prediction model with characteristic band as input. The results show that the prediction model of feature wavelength screening by GA has achieved better results. The prediction evaluation indexes of GA-MLR model are shown in the article. Table 5 summarizes the application of fruit quality detection based on phenotypic information through hyperspectral images in recent years.

Through Table 5, it can be seen that the acquisition environment of fruit hyperspectral images required by researchers still stays in a relatively stable acquisition environment such as laboratory and closed dark box. Such an environment does not require strong mobility of imaging equipment, and significantly reduces the impact of external light sources on images.

3.2 Fruit disease damage detection based on hyperspectral image

Many researchers have focused on the use of hyperspectral imaging technology to detect fruit phenotype information to achieve damage detection for fruit diseases, including disease identification, classification and quantification (Du et al., 2020). Compared with traditional disease damage detection, hyperspectral imaging technology provides an efficient non-destructive detection method (Wang et al., 2015). This section mainly introduces the application of hyperspectral image in prunoideae fruit disease damage detection.

The types of fruit disease damage are diverse and complex. Due to the existence of disease damage, the phenotypic information expressed by the fruit is abnormal, which reflects the abnormal spectral characteristics on the spectral image. In 2016, Li et al. completed the detection and recognition of nine defects in a two-color peach based on hyperspectral images (Li et al., 2016). The wavelength range of the hyperspectral image is 325-1100 nm, the spectral resolution is 2.8 nm, the distance between the sample and the lens is set to 400 mm, and the image acquisition environment is a closed darkroom. Principal component regression and PCA was used to complete the wavelength selection of hyperspectral image. In the study, according to the weight coefficient of each band, the spectral wavelength image corresponding to the maximum difference of weight coefficient is selected to obtain the dual wavelength ratio image and complete the detection of disease damage. The results show that the defect recognition effect of ratio image is better than that of single wavelength image, and the total detection rate of nine defects is 96.6%. In the subsequent study, the Li 's team completed the detection of early bruises on peaches (Li et al., 2018c). Accurate non-destructive testing of early bruises is a challenging task. Due to being in the early stage of bruises, peaches have not significantly improved in appearance. In the study, they used two hyperspectral spectrometers for image acquisition, with spectral ranges of 325-1100 nm and 930-2548 nm, respectively, and used PCA to complete the wavelength screening of hyperspectral images. They proposed an improved watershed segmentation algorithm and compared it with the traditional Ostu segmentation algorithm and global threshold segmentation method. The results of the study showed that the accurate recognition rate of damaged peaches reached 96.5%, and the accurate recognition rate of healthy peaches reached 97.5%. Their proposed method of using hyperspectral images combined with improved watershed segmentation algorithm to detect early peach bruises is an effective method that can accurately and non-destructively detect early peach bruises.

Peach is a perishable fruit, in order to extend the storage time of peaches, peaches are usually stored at low temperature. Peaches will suffer from freezing damage due to low temperature for a long time, resulting in a decrease in their market value. The symptoms are peach fruit texture deterioration and lack of juice. Pan et al. established a hyperspectral imaging system based on laboratory environment to detect cold injury of peach (Pan et al., 2016). The wavelength range of hyperspectral image used was 400-1000 nm, and the lens was 400 mm above the sample. The discrimination model of peach fruit cold injury was established by using multi-layer perception artificial neural network (MLPANN). The comparison test of full spectrum input and eight optimal wavelengths input was completed. The results showed that the overall classification accuracy of MLPANN

TABLE 5 Summary of fruit quality detection based on phenotypic information from hyperspectral imaging.

Research Object	Biochemical parameter	Imaging environment	Wavelength coverage	Wavelength screening	Model	Evaluating indicator	Ref.
Peach slices	Protopectin, Water-soluble pectin and Total pectin	Laboratory environment	380-1030 nm and 874-1734 nm	SPA, UVE, CARS	PLSR, LS-SVM	Protopectin: RPD=2.264; Water-soluble pectin and Total pectin are poor predictors	(Zhu et al., 2017)
Peach	SSC	Sealed black box	325-1100 nm	MC-UVE, CARS, RF	PLS	Set-I: r_p=0.9192, RMSEP=0.3967 Set-III: r_p=0.8469, RMSEP=0.4260	(Li et al., 2018d)
Plum	Color, firmness, SSC	Laboratory environment	600-975 nm and 865-1610 nm	PCA	PLSR	SSC: r_p^2>0.8 L* and a*: r_p^2>0.7	(Li et al., 2018a)
Green plun	SSC	Laboratory environment	400-1000 nm	SAE	BP, SVR, PLSR, SAE-BP, SAE-SVP, SAE-PLSR	SAE-PLSR: r_p=0.938; RMSEP=0.654	(Shen et al., 2020)
Peach	SSC	Sealed black box	900-1740 nm	Stacked Autoencoder	random forest	Calibration set R^2 = 0.9184; Validation set R^2 = 0.8838	(Yang et al., 2020)
Cherry	TSS and firmness	Dark room	500-1500 nm	VIP	PLSR, GPR	GPR: TSS: RPD_T=3.4, R^2_T=0.88, $RMSE_T$=0.43%; Firmness: RPD_T=2.54, R^2_T=0.60, $RMSE_T$=0.38N; PICP: 0.90-0.97	(Pullanagari and Li, 2020)
Peach	SSC, firmness, diameter, weight	Laboratory environment	400-1000 nm	CARS and RF	CARS-MLR; MLR	SSC: R^2_V: 0.841, RMSEV: 0.546; RPD: 2.51, Firmness: R^2_V: 0.826; $RMSE_V$: 1.008, RPD: 2.401 Weight: R^2_v=0.957, $RMSE_V$: 9.203, RPD: 4.819	(Xuan et al., 2022)
Cherry	SSC and PH	Sealed black box	874-1734 nm	SPR; GA	PCA; PLSR; GA-MLR; SPA-MLR	GA-MLR: SSC: R_C^2 = 0.897, RMSEC= 1.054, R_P^2 = 0.863, RMSEP=1.210 RPD=2.700	(Li et al., 2018e)

peach cold injury discriminant model based on characteristic wavelength was 92.9%. Pan et al. also proposed the cold injury (CI) index to quantify the frostbite degree of peach fruit. The formula for obtaining the cold injury (CI) index is as follows: CI index = [(CI score) × (number of fruits with this CI score)]/(3 × total number of fruits). Sun, a member of the same team, studied chlorophyll content to detect decayed honey peaches in the following year (Sun et al., 2017). In this study, the wavelength range of hyperspectral imaging is 400-1000 nm, and the sample is 22 cm closer to the lens. For the spectral data of hyperspectral images, the successive projection algorithm (SPA) is used to complete the selection of characteristic wavelengths (617,675,818 nm). The prediction models of chlorophyll content based on PLS and SPA-PLS were established. The prediction sets r_p of the two models were 0.904 and 0.858, respectively, and RMSEP were 0.633% and 0.751%, respectively. Based on the characteristic wavelength ratio image, the recognition of peach decay area was completed. The results showed that the classification accuracy of healthy peach and diseased peach was 98.75% by using three characteristic wavelength ratio images. In the same year, Sun developed a hyperspectral imaging system with a 360° rotating platform to detect varying degrees of fungal infection in peaches (Sun et al., 2018). Such improvements are very helpful for the identification of decay occurring in different locations of peach fruit. In the study, the disease damage area was quantified into four categories, no decay, mild decay, moderate decay, and severe decay, with recognition rates of 95%, 66.29%, 100%, and 100%, respectively. The low recognition of mild decay can be attributed to the fact that the changes of internal physical and chemical properties of mild decay peaches are less affected by fungal infection, so they are not easy to be detected and identified.

Li et al. 's research on peach disease damage classification was done from the perspective of time (Li et al., 2021). The spectral reflectance of bruised fruits will decrease due to the presence of bruises, which is related to the firmness, density, titratable acid (TA), vitamin C and moisture after bruises. The reflectance of fruits in different time periods after bruises also changes, and as time increases, the color of the bruise position will gradually brown. Based on this, Li et al. used hyperspectral imaging technology to complete the classification of peach bruises at different time periods (12h, 24h, 36h, 48h). The

wavelength range of hyperspectral imaging technology is 380-1080 nm, the spectral resolution is 2.8 nm, and the imaging environment is a closed darkroom environment. In the study, PCA was used to complete the screening of spectral characteristic wavelengths, and the average gray value feature of the bruise position was used as the image feature. A discriminant classification model based on PLS-DA and LS-SVM was established. The results showed that when the input variables were spectral features, the classification accuracy of PLS-DA model for peaches at 12,24,36 and 48 h after bruising was 96.67%, 96.67%, 93.33% and 83.33% respectively, and the correlation coefficient of training set was $r_c = 0.928$. Based on the LS-SVM algorithm, the correct classification of 12,24,36 and 48 h after bruising was 80%, 96.67%, 100% and 100%. This study can complete the detection of bruises as soon as possible. When the damage is softened and visible to the naked eye, the peach has lost its market value. Early detection of damage is necessary to reduce the loss. Table 6 summarizes the research on fruit disease damage detection based on phenotypic information through hyperspectral images in recent years.

Through the summary of Table 6, we found that the types of disease damage detected based on hyperspectral images are abundant, including different types of defects located on the surface and inside. It is one of the advantages of hyperspectral imaging technology that the detection of multiple types of defects in the interior of the fruit is completed by the change of spectral characteristics presented by the acquired hyperspectral image, thus ensuring the integrity of the fruit. At the same time, we can also see that the detection objects in the reviewed related studies are concentrated on peaches. This may be because the volume of peaches is easier to complete the related research in the fruit of Prunoideae, and the peach fruit is more conducive to the progress of related research due to the low light reflection phenomenon on the surface.

3.3 Other related research based on fruit hyperspectral image

As a new detection technology, hyperspectral imaging technology is widely used by researchers because of its high efficiency, non-destructive and accurate detection characteristics. In addition to the above summary of the article, there are other applications related reports.

Munera 's team used hyperspectral images to identify nectarine varieties with similar appearance but different varieties (Munera et al., 2018). The imaging method of hyperspectral image is reflection imaging, and the wavelength range is 450-1040 nm. In the study, it was found that there were significant differences in the spectra of the two nectarines in the wavelength range near 680 nm and 970 nm, which indicated that there were differences in chlorophyll (Herrero-Langreo et al., 2011) and water content (Lu and Peng, 2006) between the two nectarines. The regression coefficient vector was used to determine the 14 optimal wavelengths, and a variety classification model based on PLS-DA was established. The average spectrum of a single fruit was used as the model input feature. The results show that the classification accuracy reaches 96.3%. The study also compared the variety classification based on hyperspectral images and color visual images. The results

TABLE 6 Summary for the detection of fruit disease and damage based on phenotypic information using hyperspectral imaging.

Research object	Disease damage type	Wavelength range (nm)	Feature extraction	Method	Result	Ref.
Bi-colored peaches	Skin injury, scarring, insect, damage, puncture, injury, decay, disease, spots, dehiscent, scarring and anthracnose	325-1100 nm	PCA	Ratio images	Accuracy: 96.6%	(Li et al., 2016)
Peach	Early abrasions	325-1100 nm and 930-2548 nm	PCA	Image segmentation	Damage peach: 96.5% Non-destructive peach: 97.5%	(Li et al., 2018c)
Peach	Cold injury	400-1000 nm	/	MLP, ANN	Accuracy of cold-injured peaches: 95.8%	(Pan et al., 2016)
Honey peaches	Peach decay	400-1000 nm	SPA	PLS, SPA-PLS	PLS: r_p=0.904, RMSEP=0.751% SPA-PLS: r_p=0.858, RMSEP=0.633%	(Sun et al., 2017)
Peach	Fungal infection	400-1000 nm	SPA	PLS-DA	No decay: 95% Slight decay: 66.29% Moderate decay: 100% Severe decay: 100%	(Sun et al., 2018)
Peach	Bruises	380-1080 nm	PCA	PLS-DA, LS-SVM	12h, 24h, 36h, 48h after bruising; PLS-DA: 96.67%, 96.67%, 93.33% and 83.33% LS-SVM: 80%, 96.67%, 100%, 100%	(Li et al., 2021)

showed that the variety discrimination ability based on color vision image was low due to the high similarity of geometry and color of two nectarine varieties. In recent years, with the rapid development of image processing technology, spectral image and analysis technology, computer and data processing technology, hyperspectral imaging technology has been widely used in fruit phenotypic information and related research, such as maturity detection, fruit quality grading, disease damage identification and so on.

The image information and rich spectral data contained in hyperspectral images are the basis for the wide application of hyperspectral technology. For different use requirements and application fields, it is also necessary to select the wavelength range with the strongest correlation. Therefore, the choice of the optimal wavelength is also one of the problems faced by researchers. With the gradual improvement of machine learning theory, wavelength screening methods are also divided into two main methods: statistical methods and machine learning methods. There are more and more literatures using machine learning methods for data dimensionality reduction and model building, indicating that the combination of fruit phenotype information research based on hyperspectral image and intelligent computing is the general trend. However, there is no uniform and clear evaluation criteria for evaluating traditional modeling methods and machine learning-based modeling methods. In future research, the promotion of advanced technologies in practical production applications is also a field that cannot be ignored, pushing imaging equipment toward smaller, more sophisticated, lower cost, wider use of the environment, and faster, more accurate, and more efficient algorithms for wavelength selection and prediction models.

4 Phenotypic information acquisition and related applications based on multispectral image

In recent years, researchers have been exploring the use of spectral imaging equipment to obtain fruit phenotypic information, and based on the obtained fruit phenotypic information to achieve fruit disease damage (Kemps et al., 2010), fruit maturity and biochemical content (Li et al., 2018d) and other related applications. Multispectral images usually contain several to more than a dozen spectral bands, and some groups contained in organic substances, such as C-H, O-H, N-H, etc., these groups absorb energy in the spectral imaging bands used by multispectral imaging devices, resulting in changes in reflection or transmission spectra (Saranwong et al., 2004).

After the light beam emitted by the light source of the multispectral imaging device irradiates the surface of the fruit to be tested, part of the light beam will be reflected back after reaching the fruit to be tested, while the other part of the light beam will penetrate the fruit and scatter in different directions. Therefore, the imaging mode of the multispectral imaging device includes reflection imaging and transmission imaging. When the content of substance to be detected (SSC, TA, etc.) of fruit samples is different, different fruit samples will produce different spectral curves. Therefore, the corresponding relationship between the spectral data and the substance to be detected can be established according to the spectral characteristics of fruit samples, so as to realize the quantitative analysis of the content of the detected substance (Pissard et al., 2013). The researchers completed the measurement and analysis of relevant biomass, and combined the fruit biomass measurement research with practical application to explore the fruit maturity, disease damage, biochemical content and other related studies based on phenotype information.

4.1 Fruit maturity and biochemical parameters detection based on multispectral image

There are many criteria for determining the fruit maturity. These indicators include phenotypic information such as physical properties of the fruit (Herrero-Langreo et al., 2012) and biochemical parameters (Herrero-Langreo et al., 2011).Accurately grasping the fruit ripening stage is of great significance for guiding fruit picking time, post-harvest storage and fruit flavor. In recent years, significant progress has been made in the research on the acquisition of Prunoideae fruit phenotypic information based on multispectral images. The relevant research literature is summarized in Table 7.

Firmness is one of the most significant indicators of change during fruit ripening. Traditional methods for measuring fruit firmness such as micro-deformation measuring instruments and Magness-Taylor penetration/deformation resistance measuring instruments. However, these methods are harmful to fruit integrity and have the disadvantages of time-consuming and low accuracy. Spanish researcher M.Ruiz-Altisent first studied the relationship between peach fruit hardness and spectral wavelength to reflect the maturity of peaches (Ruiz-Altisent et al., 2006), and designed traditional measurement methods to verify the accuracy of optical measurement of peach fruit hardness. In the study, the spectral wavelengths of 450 nm and 680 nm were finally determined. The results of principal component analysis showed that the above two wavelengths were independent and complementary. The prediction model of spectral reflectance and hardness was established, the determination coefficient R^2 = 0.78. At the same time, it was found that there were some differences in the setting of parameters when establishing prediction models for different varieties of peach fruits. The two spectral wavelengths obtained

TABLE 7 Summary table of fruit maturity and biochemical parameters detection based on multispectral image.

Research object	Wavelength coverage	Testing index	Processing method	Result	Ref.
Peaches	450, 689 nm;	Firmness	PCA+MLR	R^2 = 0.78	(Ruiz-Altisent et al., 2006)
Peach	632, 650, 670, 780, 850, 900 nm	SSC and firmness	Lorentzian distribution (LD), Gaussian distribution (GD), Exponential distribution (ED) MLR	r=0.949, SEP=1.56 N, r=0.970, SEP=0.69°Brix	(Muhua et al., 2007)
Peach	450, 675, 800 nm	Firmness	Ward classification algorithm, ANOVA, regressions	R/IR imaged>R image	(Lleó et al., 2009)
Cherry	550, 660, 735, 790 nm	Antioxidant Activity	XGBoost, RF, SVR, MLP	RMSE=6.74; MAPE=15.06	(Karydas et al., 2020)

by principal component analysis are consistent with previous studies (Lu and Peng, 2006). In the study of Lu and Peng, it was found that the reflectance at 450 nm and 680 nm was related to carotenoid and chlorophyll content, respectively.

The researchers completed a test of maturity classification program based on multispectral images of peach fruits (Herrero-Langreo et al., 2011). The multispectral image data of peach fruits used in the verification were obtained from the same variety but different years. At the same time, the fruit segmentation problem of peach fruit multispectral image is optimized, and the triangle threshold segmentation algorithm is used to complete the segmentation of the region of interest of peach fruit infrared image. Compared with Otsu method, the triangle threshold segmentation method reduces the influence of highlight spots on fruits and irregularity in image background. Comparing the segmented images using the Otsu method and the triangle threshold segmentation method, in the experiment of peach fruit size estimation, the explained variance between the fruit measurement size and the image estimation size increased from 90% (Otsu) to 96%.

In the process of fruit ripening, the most intuitive and easy to observe phenomenon is that with the change of fruit maturity, fruit hardness will decrease significantly. In-depth exploration of the reasons we can know that this is due to the decomposition of certain biochemical substances within the fruit as the maturity changes. Therefore, in the detection of fruit maturity, the researchers in addition to the detection of hardness index, fruit chlorophyll content, fruit total soluble solids content (SSC), acidity, antioxidant components and other biochemical substances were also tried to measure. The traditional method for measuring the content of biochemical substances in fruit has the disadvantages of time consuming, damaging the integrity of fruit and large error. With the continuous exploration of researchers and technological progress, the detection technology based on various types of images for fruit internal biochemical content has gradually demonstrated the potential to replace traditional methods.

In addition to using multispectral images, some multispectral indices can also be used to detect fruit maturity. In subsequent reports, the researchers used four spectral indicators to detect peach fruit maturity (Lleó et al., 2011). It includes two new optical indices Ind1 and Ind2, and two previously used indices Ind3 (Sims and Gamon, 2002; Lleó et al., 2009) and IAD (Ziosi et al., 2008). The selection of the four optical indexes is within the range of chlorophyll absorption peak, and the change of optical index reflects the change of chlorophyll in fruit with maturity. The ability of four spectral indicators to distinguish fruit maturity was evaluated from two perspectives: (1) Maturity perception. The parameter Λ is introduced to evaluate the ability of each optical index to distinguish the maturity of peach fruit. The Λ parameter scores of the four indicators are as follows: Λ (Ind2) > Λ (Ind1) > Λ (IAD) > Λ (Ind3). (2) Robustness of indicators related to fruit convexity. The purpose of this comparison is to analyze which indicators are affected by fruit convexity. In the test of the effect of fruit convexity on spectral indexes, the results showed that only Ind1 was affected. In the study of L. Lleó, the intensity distribution of four optical indices in peach fruit images was obtained. It was found that the index Ind2 had the highest ability to distinguish maturity and was not affected by fruit convexity. Ind2 also allows the division of mature regions in the fruit and shows the evolution of these regions during ripening. In recent reports, researchers have tried to use changes in chlorophyll absorption index I_{AD} to reflect peach fruit maturity (Ziosi et al., 2008). I_{AD} as an optical index reflects the absorbance difference between 670 nm and 720 nm wavelengths. In the experiment, the linear regression method was used to establish the relationship between I_{AD} index and chlorophyll content, appearance color (L^*, a^*, b^*), hardness, extractable juice and SSC/TA ratio of six different varieties of peach fruit. The results showed that the higher the I_{AD} value, the higher chlorophyll content, hardness, TA and b^* values, and the lower a^* value. I_{AD} was significantly positively correlated with chlorophyll ($r^2 > 0.8$) and hardness ($r^2 > 0.6$). The higher the I_{AD} value, the higher the TA content of the peach, while the SSC did not change, so the peach fruits with different I_{AD} values had different SSC/TA ratios. The results showed that I_{AD} index could divide peaches into different maturity groups according to chlorophyll content, SSC/TA and fruit firmness. They also found that such predictions were inaccurate for

the internal quality of peach fruit, which was caused by the attenuation of light in peach fruit.

The imaging systems of the above researchers are mostly laboratory multispectral imaging systems. In the study of Karydas et al., multispectral imaging equipment was installed on an unmanned aerial vehicle platform to obtain four bands of cherry orchard aerial multispectral images (550,660,735,790 nm) (Karydas et al., 2020). The detection model of antioxidant components in cherry fruit was established by machine learning method, and the free radical scavenging activity (DPPH) of cherry fruit was analyzed. Three spectral indices were extracted based on multispectral images: normalized difference vegetation index (NDVI), carotenoid reflectance index 2 (CRI2) and anthocyanin reflectance index (ARI). Four machine learning algorithms are tested: extreme gradient boosting (XGBoost), random forest (RF), support vector regression (SVR) and multi-perceptron (MLP). The smaller RMSE and mean absolute percentage error (MAPE) obtained using the XGBoost algorithm in the study of the data obtained in 2018 were 6.74 and 15.06, respectively. In further studies, Karydas et al. were able to extend the prediction of DPPH for cherries throughout the orchard to accurately predict the maturity of cherries in different regions, guiding managers in the harvesting of cherries. The spectral indices (NDVI, CRI2, ARI) used in his research are the fusion of two or more bands of spectral data obtained by certain mathematical calculations (Crocombe, 2018). With the help of similar vegetation index (VI), the relevant research conclusions can be more obvious.

Multispectral imaging technology has been proved to be a feasible method for detecting fruit maturity and biochemical parameters. Compared with the traditional method, the detection method based on multispectral image has the advantages of high efficiency and non-destructive. However, the laboratory multispectral imaging equipment has higher requirements on the use environment, which also limits the application of multispectral image detection maturity and biochemical parameters in actual production. With the development of technology, such as UAV platform multispectral imaging equipment, portable multispectral equipment has been gradually developed, compared to the laboratory imaging equipment is small and easy to use features are more significant. Very useful for non-researchers (Li et al., 2018b). The research on fruit firmness, fruit SSC, size estimation and antioxidant content based on fruit multispectral images can provide guidance for fruit harvest time, fruit quality classification and fruit postharvest storage. It improves the quality of fruit, ensures its market value, and conforms to the development concept of precision agriculture.

4.2 Fruit disease damage detection based on multispectral image

Fruit disease damage is one of the most direct factors affecting fruit quality (Sun et al., 2019). The types of disease damage can be classified as existing in the fruit surface and fruit interior. The causes of diseases exist in many processes such as fruit growth, picking, and storage. Fruits with disease damage are more likely to rot, and if not treated in time, they can even cause lesions in other normal fruits. The multispectral image of fruit contains rich phenotypic information, and the detection of disease damage by multispectral image of fruit has been verified by many researchers.

Based on the phenomenon that the spectral characteristics of different disease damage tissues in fruit multispectral images are different, the detection of various types of fruit damage can be completed by using multispectral images. An improved enhanced GA-ANN is used to detect and classify different defect types in cherry multispectral images (Guyer and Yang, 2000). In this study, multispectral images of cherries were obtained by a non-portable multispectral imager and three spectral wavelengths of 680, 920 and 1120 nm were finally selected to complete the classification of seven different tissue types (dry crack, decay, mold, good tissue, background, stem or highlight). Researcher Daniel Guyer combines GA with ANN, and GA was used to optimize the weight of multi-layer feedforward artificial neural network. The results showed that the correct recognition rate of different tissue types reached 92%, while the correct recognition and accurate quantification of tissue types was only 72%. As can be seen from the results, more errors were made in the quantification process due to the similarity of the two types of defects, some moldy tissues were mistaken for rotten tissues. Sun et al. established a multispectral structured illumination reflection imaging system for the detection of early fungal infection in peach (Sun et al., 2019). The recognition rate of early fungal infection in peach fruit reached 98.6%, and the recognition rate of early infection without disease symptoms reached 97.6%. This study is the first time that the multispectral structured light reflection imaging system has been applied to detect fungal infection in peach fruits. Images of seven wavelengths between 690 and 810 nm at three different spatial frequencies of 60, 100 and 150 m^{-1} were obtained by the structured illumination reflection imaging system, followed by demodulation to obtain alternating component (AC) images and direct component (DC) images. Based on the acquired AC image, DC image and ratio image, three image classification methods, watershed algorithm, partial least squares discriminant analysis (PLS-DA) and CNN, were used to complete the detection and recognition of peach fruit lesions. The results show that AC images with wavelength and spatial frequency of 730 nm and 100 m^{-1} have high consistency, high detection rate and accuracy in disease region recognition and region estimation. In the horizontal comparison of classification algorithms, CNN is the best, followed by watershed segmentation algorithm.

Using multi-spectral imaging technology to detect fruit damage can efficiently and accurately complete the detection requirements, and give objective and quantitative evaluation

(Alfatni et al., 2013). Therefore, fruit disease damage detection based on multi-spectral imaging technology has important research significance, and it is also very important to promote the automation, digitization and intelligent construction of fruit detection system. With the improvement of ANN algorithm, it has been widely used in image classification applications for disease damage types. The traditional image classifier needs to select feature vectors for the classification of fruit disease damage. If the feature selection is improper or insufficient, the classification accuracy will be directly affected. In comparison, the advantage of CNN model for image recognition is that its multi-level mechanism can extract and identify complex visual features, so the advantage of CNN is obvious. In addition, the defective tissues of fruits are often a combination of several types of defects, which complicates the identification and description of defects. Therefore, for each type of defect, the combination of the selected optimal wavelength and image processing operations can help to better complete the identification and detection. The wavelength composition of the fruit multispectral image is shown in Figure 3.

5 Phenotypic information acquisition and related applications based on other types of image

In the summary and review of the acquisition of phenotypic information of Prunoideae fruits based on multispectral images, hyperspectral images and RGB images and related applications, we also found that some other imaging techniques were used in the related research of phenotypic information acquisition of Prunoideae fruits, including thermal imaging technology, computer tomography (CT) technology, laser-light backscattering imaging (LLBI) and so on. These imaging techniques have not been widely used in the related research

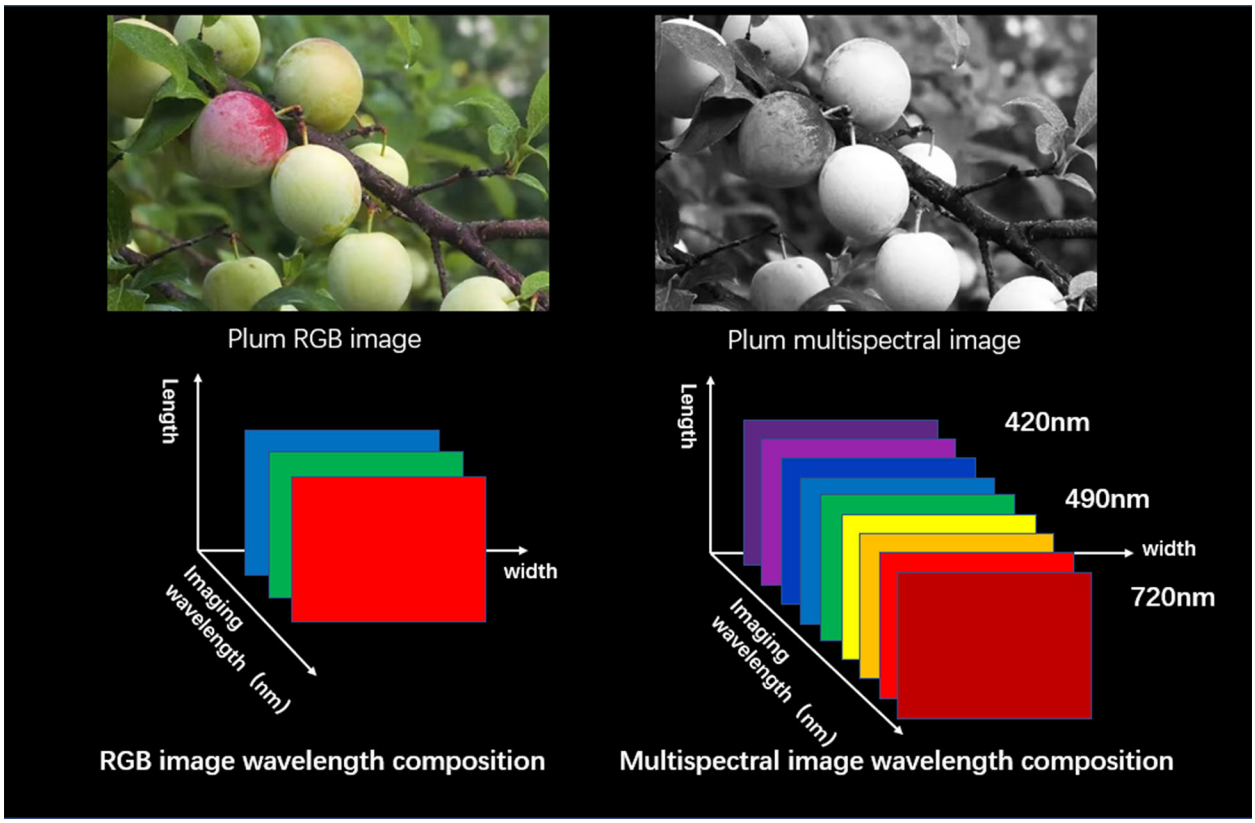


FIGURE 3
Wavelength composition of multispectral image.

of the Prunoideae fruit, but they cannot be ignored. In the following sections, we analyze and review the cited literature according to different imaging techniques.

5.1 Phenotypic information acquisition based on thermal imaging technology and related research

The thermal imaging equipment uses the infrared detector and the optical imaging mirror to receive the infrared radiation energy of the measured target, and reflects the energy distribution to the photosensitive element of the infrared detector, so as to obtain the thermal image of the measured target. The obtained thermal image corresponds to the thermal distribution of the measured object (Lee et al., 2019). The imaging methods of thermal imaging equipment include active system and passive system. Compared with passive system, active system usually includes heating or cooling system.

Cherry is a kind of fruit which is sensitive to the surface temperature and humidity. Fruit cracking caused by abnormal changes in fruit surface temperature and humidity will seriously affect its market value. Osroosh et al. used thermal imager-RGB image system to detect the surface temperature of cherry fruit and used microclimate information detection system to verify the results (Osroosh and Peters, 2019). The thermal imaging equipment was installed at a distance of 20 cm from the target cherry. In the experiment, a simulated rainfall system was also built to change the surface humidity of the cherry fruit. They also developed a custom computer vision algorithm to recognize cherry fruits in thermal images and RGB images, and completed the extraction of fruit surface temperature. In the study, they found that the surface temperature of cherry fruits was highly correlated with the surface temperature of leaves ($R^2 > 0.89$). The final experimental results show that it is feasible to detect the surface temperature of cherry fruit using a system based on low-resolution thermal RGB images. They also established the normalized temperature index (NTDI and NRTI) to quantify fruit surface humidity levels.

In a recently reported study, researchers developed two models for predicting cherry surface humidity based on thermal-RGB images and weather sensing systems (Ranjan et al., 2022). The input data of the first model is weather sensor data, and the input data of the other model combines the fruit surface temperature obtained from thermal image data. An automatic custom image processing algorithm was developed for fruit recognition and surface temperature extraction, and the radiation calibration equation was used to correct the temperature data. In the experiment, two varieties of cherry fruits were used to complete the prediction of fruit surface humidity. The results showed that the correlation between the measured and predicted values of humidity was $R^2 = 0.80$ and $R^2 = 0.86$, respectively. Compared with other imaging technologies, thermal imaging technology can accurately provide the temperature information of the measured target, which is an advantage that other types of images do not have. Especially in the exploration of temperature and fruit phenotype research applications, has an irreplaceable role. However, the use of thermal imaging equipment in outdoor environments is strongly affected by light, which will have a certain impact on its accuracy.

5.2 Phenotypic information acquisition based on computed tomography imaging technology and related research

Computer tomography (CT) has become one of the mature non-destructive technologies for measuring the external morphological characteristics and internal defect detection of agricultural products (Arendse et al., 2018). The CT technology ray includes X-ray, γ-ray, ultrasound, etc. CT technology can reflect the density change of the measured sample. The density and absorption coefficient of the measured sample will lead to the attenuation of the ray during the penetration process. The CT image can reflect the density change of the measured object through the gray value of the pixel. The image is white, indicating high density, and black indicates low density.

Kritzinger et al. used X-ray CT technology to detect the occurrence of fruit core cracking during plum fruit development (Kritzinger et al., 2017). They selected six plum varieties to explore the causes of fruit core cracking, and randomly selected measurement targets in the orchard for X-ray CT scanning. The acquisition process of CT images developed from the inner epidermis until the fruit core was completely hardened. The results showed that due to the influence of temperature changes, the growth of the inner epidermis of the fruit was affected, and the incompletely hardened inner epidermis was affected by the tension of the pulp, resulting in cracking. In their research, CT technology accurately reflected the process of fruit core hardening at different stages, thus accurately discovering the occurrence of fruit core cracking.

In a study to determine whether 1-Methylcyclopropene treatment of apricot fruit is beneficial for fruit preservation, X-ray CT technology was used to detect the occurrence of voids inside the fruit (Karmoker et al., 2018). The obtained fruit CT images are divided into high-density regions (−200 ~ +350HU) and low-density regions (−900 ~ −200HU) to detect voids. The changes of ethylene content in fruits treated with conventional methods and 1-Methylcyclopropene during storage were compared. The results showed that the apricot fruit pretreated with 1-Methylcyclopropene effectively inhibited the production of ethylene. From the CT image of the fruit, it can be seen that the inside voids the treated apricot fruit is less than that of the

apricot fruit preserved by the conventional method. These results show that CT imaging technology has significant advantages in the detection of fruit internal.

5.3 Phenotypic information acquisition based on laser-light backscattering imaging technology and related research

Laser-light backscattering imaging (LLBI), as a low-cost imaging technology, realizes the detection and analysis of targets by using the principles of light absorption, scattering and image processing in the visible and near-infrared electromagnetic spectrum. When a beam of bright light is irradiated to the fruit surface, most of the light will transmit to the fruit tissue, and the other part will diffuse to the fruit surface (Adebayo et al., 2016). Through the interaction between light and the object to be measured, useful information about the structure and composition of the object to be measured is provided for quality analysis of fruits.

In 2016, a quality evaluation study using LLBI technology to detect plum fruit during tree development and storage was reported (Rezaei Kalaj et al., 2016). In this study, two different wavelengths of light (532,785nm) were used to detect two different varieties of plum. The full width at half maximum (FWHM) value was obtained by radial backscattering profile calculation, and this value was tried to establish with fruit hardness, SSC, dry matter content and normalized anthocyanin index. The results showed that the decrease of FWHM532 was closely related to the increase of anthocyanin content during fruit development. In addition, the increase of FWHM785 was closely related to the decrease of flesh firmness during fruit development and storage. Their results show that it is feasible to use the appropriate wavelength LLBI technique to nondestructively detect the oxidation resistance and firmness of plum fruit.

In a recent study, LLBI technique was used to predict the quality characteristics and maturity of apricot fruit (Mozaffari et al., 2022). Different from previous studies, the wavelength of 650 nm was used to obtain backscatter images of six ripening stages of apricot fruit. They used Otsu and first inflection point techniques to segment the image and extract spatial domain features from it. They established three prediction models (ANN, PLSR, PCA-ANN) to predict the hardness and SSC of apricot. The results show that the R^2 based on ANN prediction model is the highest and RMSE is the lowest. The results of hardness and SSC were $R^2_{CV} = 0.974$, $RMSE_{CV} = 3.482$ and $R^2_{CV} = 0.963$, $RMSE_{CV} = 1.146$, respectively.

6 Discussion

At present, the acquisition of fruit phenotypic information based on image technology and its related research have made remarkable progress under the efforts of researchers. In the review of past research, our research based on image type is divided into three types. Here, we discuss the related research of three image types.

(1) Fruit phenotypic information acquisition and related research based on RGB images: Firstly, from the perspective of images, RGB images only contain image information compared with the former two, and do not involve spectral data. In terms of phenotypic information acquisition, only fruit color features, texture features and geometric morphological parameters based on image information can be extracted. However, the convenience of RGB image acquisition is one of its most significant advantages.

The convenience of RGB image acquisition is one of its most significant advantages. However, with the popularity of various imaging devices, the obtained RGB images have differences in pixel resolution, image size, etc. These differences should be solved in subsequent image processing. In addition, the acquisition of RGB images should also take into account the automatic preprocessing of the image during the acquisition of the image by the smartphone, such as the adjustment of exposure time and contrast, which will lead to the acquisition of RGB images can not accurately reflect the color characteristics of the target object.

(2) Fruit phenotypic information acquisition and related research based on hyperspectral images: Compared with multispectral images, hyperspectral images have become one of the research hotspots with richer spectral data. The spectral resolution of hyperspectral imaging equipment is nanoscale, and the hyperspectral image obtained often contains hundreds of thousands of wavelengths. Therefore, the rich spectral data makes the hyperspectral image more comprehensive and accurate in the process of obtaining fruit phenotypic information, which makes it widely studied in the fields of fruit maturity detection, fruit quality grading, disease damage identification quantification and so on.

Removing redundant hyperspectral data is also one of the hot issues for researchers in related research using hyperspectral images. With the improvement of artificial intelligence theory, there is also great potential in feature wavelength selection algorithms, such as UVE, SVE, etc., which have achieved good results. It should be mentioned that in most of the studies reviewed, hyperspectral imaging equipment is used in a laboratory environment, and some researchers have also used portable spectral imaging equipment for research. However, further improvement is needed to balance image quality and cost issues.

(3) Fruit phenotypic information acquisition and related research based on multispectral images: Fruit multispectral images are used by researchers in fruit maturity, biochemical parameter detection, disease damage detection and other fields by virtue of the image information and spectral data contained in the image. On the one hand, the band range of multispectral images

often contains only a few to a dozen, and there is a defect that the spectral data acquisition is not comprehensive, which limits the researchers to obtain various phenotypic information of fruits through spectral data and complete related applications. On the other hand, the low number of spectral data in multispectral images also reduces the complexity of data dimensionality reduction and further processing in the later stage. From the perspective of multispectral imaging equipment, the low cost and portability of multispectral imaging equipment are its most significant features. Considering that the imaging environment is mostly outdoor, sunlight will have a certain impact on multispectral images, so multispectral imaging equipment should be further improved in terms of resolution and anti-interference ability.

(4) In the review and summary of other imaging technology, we mentioned thermal imaging technology, CT imaging technology, LLBI technology. The images obtained by these imaging techniques can provide data in different dimensions for the acquisition of fruit phenotypic information and subsequent research applications. The acquisition of fruit surface temperature information by thermal imaging technology is irreplaceable in the study of avoiding fruit cracking due to abnormal surface temperature. CT imaging technology and LLBI technology can accurately reflect the internal quality of fruits by means of X-ray and laser beam penetration measurement in the detection of internal phenotypic information of fruits.

Finally, with the development of image processing technology and related target recognition and segmentation algorithm, the acquisition of fruit phenotype information based on RGB image and related research is still one of the important areas. Researchers have used a variety of neural network algorithms to complete research on maturity, fruit quality, disease damage and other related research based on RGB images. In this review article, we do not deeply explore the related algorithms used by researchers, but it can be concluded that with the continuous improvement of related algorithms, deep learning shows significant advantages over traditional algorithms.

7 Conclusions

We tried to review the acquisition of fruit phenotypic information and related research based on image technology. Image types mainly include multispectral, hyperspectral and RGB images. Then, according to the research purpose, the research based on each type of image is introduced, including the imaging equipment parameters, image analysis technology and research results, etc. When the actual design is based on different types of images to complete the application of fruit detection, due to the differences, it may be necessary to readjust the various parameters of the image analysis algorithm and combine various technologies. In addition, the specificity of the imaging equipment and imaging environment used in the references leads to its popularization to be improved. In some researches based on natural light and outdoor environments, the preprocessing algorithms proposed for other interference light sources are worthy of reference, which can significantly reduce the impact of noise in target fruit recognition and detection, thereby improving the accuracy of research. In order to further promote the application of image-based fruit detection technology in actual production, it is necessary to further improve the portability of imaging equipment and coordinate the cost problem. Moreover, it can be seen from the environmental specificity of previous studies that the existing analysis algorithms have poor universality. Therefore, the development of new algorithms to achieve high efficiency, high accuracy and strong adaptability is of great significance for the promotion of this technology in practical applications.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

XL collected and analyzed references. XL and NL prepared the review and wrote the manuscript; YH and XJL offered much help in the process of revision. ZR provided financial supports. All authors read and approved the final manuscript.

Funding

The authors are grateful for the financial support from the Hebei Provincial Department of Science and Technology (Grant number 22327203D).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adebayo, S. E., Hashim, N., Abdan, K., and Hanafi, M. (2016). Application and potential of backscattering imaging techniques in agricultural and food processing - a review. *J. Food Eng.* 169, 155–164. doi: 10.1016/j.jfoodeng.2015.08.006

Ahmad, J., Jan, B., Farman, H., Ahmad, W., and Ullah, A. (2020). Disease detection in plum using convolutional neural network under true field conditions. *Sensors (Switzerland)* 20, 1–18. doi: 10.3390/s20195569

Ahmad, J., Muhammad, K., Ahmad, I., Ahmad, W., Smith, M. L., Smith, L. N., et al. (2018a). Visual features based boosted classification of weeds for real-time selective herbicide sprayer systems. *Comput. Ind.* 98, 23–33. doi: 10.1016/j.compind.2018.02.005

Ahmad, J., Muhammad, K., Lee, M. Y., and Baik, S. W. (2017). Endoscopic image classification and retrieval using clustered convolutional features. *J. Med. Syst.* 41, 196–208. doi: 10.1007/s10916-017-0836-y

Ahmad, J., Muhammad, K., Lloret, J., and Baik, S. W. (2018b). Efficient conversion of deep features to compact binary codes using Fourier decomposition for multimedia big data. *IEEE Trans. Industr Inform* 14, 3205–3215. doi: 10.1109/TII.2018.2800163

Alfatni, M. S. M., Shariff, A. R. M., Abdullah, M. Z., Marhaban, M. H. B., and Saaed, O. M. B. (2013). The application of internal grading system technologies for agricultural products - review. *J. Food Eng.* 116, 703–725. doi: 10.1016/j.jfoodeng.2013.01.001

Ali, M. M., Bachik, N. A., Muhadi, N.'A, Tuan Yusof, T. N., and Gomes, C. (2019). Non-destructive techniques of detecting plant diseases: A review. *Physiol. Mol. Plant Pathol.* 108, 1–12. doi: 10.1016/j.pmpp.2019.101426

Alosaimi, W., Alyami, H., and Uddin, M. I. (2021). PeachNet: Peach diseases detection for automatic harvesting. *Computers Materials Continua* 67, 1665–1677. doi: 10.32604/cmc.2021.014950

Arendse, E., Fawole, O. A., Magwaza, L. S., and Opara, U. L. (2018). Non-destructive prediction of internal and external quality attributes of fruit with thick rind: A review. *J. Food Eng.* 217, 11–23. doi: 10.1016/j.jfoodeng.2017.08.009

Blasco, J., Munera, S., Aleixos, N., Cubero, S., and Molto, E. (2017). "Machine vision-based measurement systems for fruit and vegetable quality control in postharvest." *Advances in Biochemical Engineering-Biotechnology* 161, 71–91. doi: 10.1007/10_2016_51

Chattaraj, S., Chakraborty, D., Garg, R. N., Singh, G. P., Gupta, V. K., Singh, S., et al. (2013). Hyperspectral remote sensing for growth-stage-specific water use in wheat. *Field Crops Res.* 144, 179–191. doi: 10.1016/j.fcr.2012.12.009

Choudhury, S. D., Samal, A., and Awada, T. (2019). Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00508

Crocombe, R. A. (2018). Portable spectroscopy. *Appl. Spectrosc* 72, 1701–1751. doi: 10.1177/0003702818809719

Dhondt, S., Wuyts, N., and Inzé, D. (2013). Cell to whole-plant phenotyping: The best is yet to come. *Trends Plant Sci.* 18, 428–439. doi: 10.1016/j.tplants.2013.04.008

Du, Z., Zeng, X., Li, X., Ding, X., Cao, J., and Jiang, W. (2020). Recent advances in imaging techniques for bruise detection in fruits and vegetables. *Trends Food Sci. Technol.* 99, 133–141. doi: 10.1016/j.tifs.2020.02.024

Esehaghbeygi, A., Ardforoushan, M., Monajemi, S. A. H., and Masoumi, A. A. (2010). Digital image processing for quality ranking of saffron peach. *Int. Agrophysics* 24, 115–120.

Fernández-Novales, J., Garde-Cerdán, T., Tardáguila, J., Gutiérrez-Gamboa, G., Pérez-Álvarez, E. P., and Diago, M. P. (2019). Assessment of amino acids and total soluble solids in intact grape berries using contactless vis and NIR spectroscopy during ripening. *Talanta* 199, 244–253. doi: 10.1016/j.talanta.2019.02.037

Fu, L., Gao, F., Wu, J., Li, R., Karkee, M., and Zhang, Q. (2020). Application of consumer RGB-d cameras for fruit detection and localization in field: A critical review. *Comput. Electron Agric.* 177. doi: 10.1016/j.compag.2020.105687

Gómez-Sanchis, J., Moltó, E., Camps-Valls, G., Gómez-Chova, L., Aleixos, N., and Blasco, J. (2008). Automatic correction of the effects of the light source on spherical objects. an application to the analysis of hyperspectral images of citrus fruits. *J. Food Eng.* 85, 191–200. doi: 10.1016/j.jfoodeng.2007.06.036

Guyer, D., and Yang, X. (2000). Use of genetic artificial neural networks and spectral imaging for defect detection on cherries. *Comput Electron Agric* 29, 179–194. doi: 10.1016/S0168-1699(00)00146-0.

Herrero-Langreo, A., Fernández-Ahumada, E., Roger, J. M., Palagós, B., and Lleó, L. (2012). Combination of optical and non-destructive mechanical techniques for the measurement of maturity in peach. *J. Food Eng.* 108, 150–157. doi: 10.1016/j.jfoodeng.2011.07.004

Herrero-Langreo, A., Lunadei, L., Lleó, L., Diezma, B., and Ruiz-Altisent, M. (2011). Multispectral vision for monitoring peach ripeness. *J. Food Sci.* 76, 178–187. doi: 10.1111/j.1750-3841.2010.02000.x

Huang, H., Liu, L., and Ngadi, M. O. (2014). Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors (Switzerland)* 14, 7248–7276. doi: 10.3390/s140407248

Huang, S., Zhou, G., He, M., Chen, A., Zhang, W., and Hu, Y. (2020). Detection of peach disease image based on asymptotic non-local means and PCNN-IPELM. *IEEE Access* 8, 136421–136433. doi: 10.1109/ACCESS.2020.3011685

Karmoker, P., Obatake, W., Tanaka, F., and Tanaka, F. (2018). Quality evaluation of 1-methylcyclopropene treated Japanese apricot using x-ray computed tomography. *Food Sci. Technol. Res.* 24, 377–384. doi: 10.3136/fstr.24.377

Karydas, C., Iatrou, M., Kouretas, D., Patouna, A., Iatrou, G., Lazos, N., et al. (2020). Prediction of antioxidant activity of cherry fruits from UAS multispectral imagery using machine learning. *Antioxidants* 9. doi: 10.3390/antiox9020156

Kaur, H., Sawhney, B. K., and Jawandha, S. K. (2018). Evaluation of plum fruit maturity by image processing techniques. *J. Food Sci. Technol.* 55, 3008–3015. doi: 10.1007/s13197-018-3220-0

Kemps, B., Leon, L., Best, S., de Baerdemaeker, J., and de Ketelaere, B. (2010). Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. *Biosyst. Eng.* 105, 507–513. doi: 10.1016/j.biosystemseng.2010.02.002

Khojastehnazhand, M., Mohammadi, V., and Minaei, S. (2019). Maturity detection and volume estimation of apricot using image processing technique. *Sci. Hortic.* 251, 247–251. doi: 10.1016/j.scienta.2019.03.033

Kopjar, M., Alilović, D., Požrl, T., Piližota, V., and Pichler, A. (2017). Phenolics content and antioxidant activity of sour cherry extracts with sugar addition. *Acta Aliment* 46, 501–507. doi: 10.1556/066.2017.46.4.13.

Kritzinger, I., Lötze, E., and Jooste, M. (2017). Stone hardening and broken stones in Japanese plums (Prunus salicina lindl.) evaluated by means of computed tomography scans. *Sci. Hortic.* 221, 1–9. doi: 10.1016/j.scienta.2017.04.008

Lee, A., Kim, S. Y., Hong, S. J., Han, Y., Choi, Y., Kim, M., et al. (2019). Phenotypic analysis of fruit crops water stress using infrared thermal imaging. *J. Biosyst. Eng.* 44, 87–94. doi: 10.1007/s42853-019-00020-2

Li, J., Chen, L., and Huang, W. (2018c). Detection of early bruises on peaches (Amygdalus persica l.) using hyperspectral imaging coupled with improved watershed segmentation algorithm. *Postharvest Biol. Technol.* 135, 104–113. doi: 10.1016/j.postharvbio.2017.09.007

Li, J., Chen, L., Huang, W., Wang, Q., Zhang, B., Tian, X., et al. (2016). Multispectral detection of skin defects of bi-colored peaches based on vis-NIR hyperspectral imaging. *Postharvest Biol. Technol.* 112, 121–133. doi: 10.1016/j.postharvbio.2015.10.007

Li, B., Cobo-Medina, M., Lecourt, J., Harrison, N. B., Harrison, R. J., and Cross, J. v. (2018a). Application of hyperspectral imaging for nondestructive measurement of plum quality attributes. *Postharvest Biol. Technol.* 141, 8–15. doi: 10.1016/j.postharvbio.2018.03.008

Li, J., Fan, S., and Huang, W. (2018d). Assessment of multiregion local models for detection of SSC of whole peach (Amygdalus persica l.) by combining both hyperspectral imaging and wavelength optimization methods. *J. Food Process Eng.* 41. doi: 10.1111/jfpe.12914

Li, X., Liu, Y., Jiang, X., and Wang, G. (2021). Supervised classification of slightly bruised peaches with respect to the time after bruising by using hyperspectral imaging technology. *Infrared Phys. Technol.* 113. doi: 10.1016/j.infrared.2020.103557

Li, X., Wei, Y., Xu, J., Feng, X., Wu, F., Zhou, R., et al. (2018e). SSC And pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology. *Postharvest Biol. Technol.* 143, 112–118. doi: 10.1016/j.postharvbio.2018.05.003

Li, C., Zhu, X., Wei, Y., Cao, S., Guo, X., Yu, X., et al. (2018b). Estimating apple tree canopy chlorophyll content based on sentinel-2A remote sensing imaging. *Sci. Rep.* 8. doi: 10.1038/s41598-018-21963-0

Lleó, L., Barreiro, P., Ruiz-Altisent, M., and Herrero, A. (2009). Multispectral images of peach related to firmness and maturity at harvest. *J. Food Eng.* 93, 229–235. doi: 10.1016/j.jfoodeng.2009.01.028

Lleó, L., Roger, J. M., Herrero-Langreo, A., Diezma-Iglesias, B., and Barreiro, P. (2011). Comparison of multispectral indexes extracted from hyperspectral images for the assessment of fruit ripening. *J. Food Eng.* 104, 612–620. doi: 10.1016/j.jfoodeng.2011.01.028

Lu, R., and Peng, Y. (2006). Hyperspectral scattering for assessing peach fruit firmness. *Biosyst. Eng.* 93, 161–171. doi: 10.1016/j.biosystemseng.2005.11.004

Mahlein, A. K. (2016). Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–254. doi: 10.1094/PDIS-03-15-0340-FE

Miragaia, R., Chávez, F., Díaz, J., Vivas, A., Prieto, M. H., and Moñino, M. J. (2021). Plum ripeness analysis in real environments using deep learning with convolutional neural networks. *Agronomy* 11, 2353. doi: 10.3390/agronomy11112353

Momeny, M., Jahanbakhshi, A., Jafarnezhad, K., and Zhang, Y. D. (2020). Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. *Postharvest Biol. Technol.* 166. doi: 10.1016/j.postharvbio.2020.111204

Mozaffari, M., Sadeghi, S., and Asefi, N. (2022). Prediction of the quality properties and maturity of apricot by laser light backscattering imaging. *Postharvest Biol. Technol.* 186. doi: 10.1016/j.postharvbio.2022.111842

Muhua, L., Peng, F., and Renfa, C. (2007). Non-destructive estimation peach SSC and firmness by mutispectral reflectance imaging. *New Z. J. Agric. Res.* 50, 601–608. doi: 10.1080/00288230709510328

Munera, S., Amigo, J. M., Aleixos, N., Talens, P., Cubero, S., and Blasco, J. (2018). Potential of VIS-NIR hyperspectral imaging and chemometric methods to identify similar cultivars of nectarine. *Food Control* 86, 1–10. doi: 10.1016/j.foodcont.2017.10.037

Munera, S., Amigo, J. M., Blasco, J., Cubero, S., Talens, P., and Aleixos, N. (2017). Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. *J. Food Eng.* 214, 29–39. doi: 10.1016/j.jfoodeng.2017.06.031

Osroosh, Y., and Peters, R. T. (2019). Detecting fruit surface wetness using a custom-built low-resolution thermal-RGB imager. *Comput. Electron Agric.* 157, 509–517. doi: 10.1016/j.compag.2019.01.023

Pan, L., Zhang, Q., Zhang, W., Sun, Y., Hu, P., and Tu, K. (2016). Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network. *Food Chem.* 192, 134–141. doi: 10.1016/j.foodchem.2015.06.106

Pissard, A., Fernández Pierna, J. A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., et al. (2013). Non-destructive measurement of vitamin c, total polyphenol and sugar content in apples using near-infrared spectroscopy. *J. Sci. Food Agric.* 93, 238–244. doi: 10.1002/jsfa.5779

Pu, Y. Y., Feng, Y. Z., and Sun, D. W. (2015). Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: A review. *Compr. Rev. Food Sci. Food Saf.* 14, 176–188. doi: 10.1111/1541-4337.12123

Pullanagari, R. R., and Li, M. (2020). Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. *J. Food Eng.* 289, 110177.1–110177.11. doi: 10.1016/j.jfoodeng.2020.110177

Qin, J., Chao, K., Kim, M. S., Lu, R., and Burks, T. F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng.* 118, 157–171. doi: 10.1016/j.jfoodeng.2013.04.001

Ranjan, R., Sinha, R., Khot, L. R., and Whiting, M. (2022). Thermal-RGB imagery and in-field weather sensing derived sweet cherry wetness prediction model. *Sci. Hortic.* 294. doi: 10.1016/j.scienta.2021.110782

Rezaei Kalaj, Y., Mollazade, K., Herppich, W., Regen, C., and Geyer, M. (2016). Changes of backscattering imaging parameter during plum fruit development on the tree and during storage. *Sci. Hortic.* 202, 63–69. doi: 10.1016/j.scienta.2016.02.029

Ropelewska, E., Popińska, W., Sabanci, K., and Aslan, M. F. (2021). Cultivar identification of sweet cherries based on texture parameters determined using image analysis. *J. Food Process Eng.* 44, e13724.1–e13724.10. doi: 10.1111/jfpe.13724

Ropelewska, E., and Rutkowski, K. P. (2021). Differentiation of peach cultivars by image analysis based on the skin, flesh, stone and seed textures. *Eur. Food Res. Technol.* 247, 2371–2377. doi: 10.1007/s00217-021-03797-9

Ruiz-Altisent, M., Lleó, L., and Riquelme, F. (2006). Instrumental quality assessment of peaches: Fusion of optical and mechanical parameters. *J. Food Eng.* 74, 490–499. doi: 10.1016/j.jfoodeng.2005.01.048

Saranwong, S., Sornsrivichai, J., and Kawano, S. (2004). Prediction of ripe-stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy. *Postharvest Biol. Technol.* 31, 137–145. doi: 10.1016/j.postharvbio.2003.08.007

Shakoor, N., Lee, S., and Mockler, T. C. (2017). High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Curr. Opin. Plant Biol.* 38, 184–192. doi: 10.1016/j.pbi.2017.05.006

Shao, Y., Xuan, G., Hu, Z., Gao, Z., and Liu, L. (2019). Determination of the bruise degree for cherry using vis-NIR reflection spectroscopy coupled with multivariate analysis. *PLoS One* 14, 1–13. doi: 10.1371/journal.pone.0222633

Shen, L., Wang, H., Liu, Y., Liu, Y., Zhang, X., and Fei, Y. (2020). Prediction of soluble solids content in green plum by using a sparse autoencoder. *Appl. Sci.* 10. doi: 10.3390/app10113769

Sims, D. A., and Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens Environ.* 81, 337–354. doi: 10.1016/S0034-4257(02)00010-X

Sohaib Ali Shah, S., Zeb, A., Qureshi, W. S., Arslan, M., Ullah Malik, A., Alasmary, W., et al. (2020). Towards fruit maturity estimation using NIR spectroscopy. *Infrared Phys. Technol.* 111, 103479. doi: 10.1016/j.infrared.2020.103479

Sun, Y., Lu, R., Lu, Y., Tu, K., and Pan, L. (2019). Detection of early decay in peaches by structured-illumination reflectance imaging. *Postharvest Biol. Technol.* 151, 68–78. doi: 10.1016/j.postharvbio.2019.01.011

Sun, Y., Wang, Y., Xiao, H., Gu, X., Pan, L., and Tu, K. (2017). Hyperspectral imaging detection of decayed honey peaches based on their chlorophyll content. *Food Chem.* 235, 194–202. doi: 10.1016/j.foodchem.2017.05.064

Sun, Y., Xiao, H., Tu, S., Sun, K., Pan, L., and Tu, K. (2018). Detecting decayed peach using a rotating hyperspectral imaging testbed. *LWT* 87, 326–332. doi: 10.1016/j.lwt.2017.08.086

Su, W.-H., and Sun, D.-W. (2018). Multispectral imaging for plant food quality analysis and visualization. *Compr. Rev. Food Sci. Food Saf.* 17, 220–239. doi: 10.1111/1541-4337.12317

Taghadomi-Saberi, S., Omid, M., Emam-Djomeh, Z., and Ahmadi, H. (2014). Evaluating the potential of artificial neural network and neuro-fuzzy techniques for estimating antioxidant activity and anthocyanin content of sweet cherry during ripening by using image processing. *J. Sci. Food Agric.* 94, 95–101. doi: 10.1002/jsfa.6202

Taghadomi-Saberi, S., Omid, M., Emam-Djomeh, Z., and Faraji-Mahyari, K. (2015). Determination of cherry color parameters during ripening by artificial neural network assisted image processing technique. *Journal of Agricultural Science and Technology* 17, 589–600. doi: 10.13140/2.1.2049.6006.

Tang, J., Deng, C., and Huang, G. (2016). Extreme learning machine for multilayer perceptron. *IEEE Trans. Neural Netw. Learn Syst.* 27, 809–821. doi: 10.1109/TNNLS.2015.2424995

Toivonen, P. M. A., Batista, A., and Lannard, B. (2017). Development of a predictive model for 'lapins' sweet cherry dry matter content using a visible/near-infrared spectrometer and its potential application to other cultivars. *Can. J. Plant Sci.* 97, 1030–1035. doi: 10.1139/cjps-2017-0013

Villacrés, J. F., and Cheein, F. A. (2020). Detection and characterization of cherries: A deep learning usability case study in Chile. *Agronomy* 10. doi: 10.3390/agronomy10060835

Wang, H., Peng, J., Xie, C., Bao, Y., and He, Y. (2015). Fruit quality evaluation using spectroscopy technology: A review. *Sensors (Switzerland)* 15, 11889–11927. doi: 10.3390/s150511889

Wang, Q., Wang, H., Xie, L., and Zhang, Q. (2012). Outdoor color rating of sweet cherries using computer vision. *Comput. Electron Agric.* 87, 113–120. doi: 10.1016/j.compag.2012.05.010

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M., and Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 667, 14–32. doi: 10.1016/j.aca.2010.03.048

Xuan, G., Gao, C., and Shao, Y. (2022). Spectral and image analysis of hyperspectral data for internal and external quality assessment of peach fruit. *Spectrochim Acta A Mol. Biomol Spectrosc* 272. doi: 10.1016/j.saa.2022.121016

Yang, B., Gao, Y., Yan, Q., Qi, L., Zhu, Y., and Wang, B. (2020). Estimation method of soluble solid content in peach based on deep features of hyperspectral imagery. *Sensors (Switzerland)* 20, 1–12. doi: 10.3390/s20185021

Yao, N., Ni, F., Wang, Z., Luo, J., Sung, W. K., Luo, C., et al. (2021). L2MXception: An improved xception network for classification of peach diseases. *Plant Methods* 17. doi: 10.1186/s13007-021-00736-3

Ye, X., Sakai, K., Garciano, L. O., Asada, S. I., and Sasao, A. (2006). Estimation of citrus yield from airborne hyperspectral images using a neural network model. *Ecol. Modell* 198, 426–432. doi: 10.1016/j.ecolmodel.2006.06.001

Yun, Y-H, Li, H-D, Leslie, L. R., Fan, W., Wang, J. J., Cao, D. S., et al. (2013). An efficient method of wavelength interval selection based on random frog for multivariate spectral calibration. *Spectrochim Acta A Mol. Biomol Spectrosc* 111, 31–36. doi: 10.1016/j.saa.2013.03.083

Zhang, B., et al. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International* 62, p. 326–343.

Zhu, N., Huang, W., Wu, D., Chen, K., and He, Y. (2017). Quantitative visualization of pectin distribution maps of peach fruits. *Sci. Rep.* 7. doi: 10.1038/s41598-017-09817-7

Zhu, N., Lin, M., Nie, Y., Wu, D., and Chen, K. (2016). Study on the quantitative measurement of firmness distribution maps at the pixel level inside peach pulp. *Comput. Electron Agric.* 130, 48–56. doi: 10.1016/j.compag.2016.09.018

Ziosi, V., Noferini, M., Fiori, G., Tadiello, A., Trainotti, L., Casadoro, G., et al. (2008). A new index based on vis spectroscopy to characterize the progression of ripening in peach fruit. *Postharvest Biol. Technol.* 49, 319–329. doi: 10.1016/j.postharvbio.2008.01.017

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 January 2023
DOI 10.3389/fpls.2023.1111175

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Wenzheng Bao,
Xuzhou University of Technology, China
Nisha Pillai,
Mississippi State University, United States

*CORRESPONDENCE
Pan Gao
gp_inf@shzu.edu.cn
Wei Xu
xuwei0412@shzu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 29 November 2022
ACCEPTED 13 January 2023
PUBLISHED 31 January 2023

CITATION
Yan J, Yan T, Ye W, Lv X, Gao P and Xu W (2023) Cotton leaf segmentation with composite backbone architecture combining convolution and attention. *Front. Plant Sci.* 14:1111175. doi: 10.3389/fpls.2023.1111175

COPYRIGHT
© 2023 Yan, Yan, Ye, Lv, Gao and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Cotton leaf segmentation with composite backbone architecture combining convolution and attention

Jingkun Yan[1,2†], Tianying Yan[1,2†], Weixin Ye[1,2], Xin Lv[2,3], Pan Gao[1,2*] and Wei Xu[3*]

[1]College of Information Science and Technology, Shihezi University, Shihezi, China, [2]National-Local Joint Engineering Research Center for Agricultural Big Data, Xinjiang Production and Construction Group, Shihezi, China, [3]College of Agriculture, Shihezi University, Shihezi, China

Plant leaf segmentation, especially leaf edge accurate recognition, is the data support for automatically measuring plant phenotypic parameters. However, adjusting the backbone in the current cutting-edge segmentation model for cotton leaf segmentation applications requires various trial and error costs (e.g., expert experience and computing costs). Thus, a simple and effective semantic segmentation architecture (our model) based on the composite backbone was proposed, considering the computational requirements of the mainstream Transformer backbone integrating attention mechanism. The composite backbone was composed of CoAtNet and Xception. CoAtNet integrated the attention mechanism of the Transformers into the convolution operation. The experimental results showed that our model outperformed the benchmark segmentation models PSPNet, DANet, CPNet, and DeepLab v3+ on the cotton leaf dataset, especially on the leaf edge segmentation (MIoU: 0.940, BIoU: 0.608). The composite backbone of our model integrated the convolution of the convolutional neural networks and the attention of the Transformers, which alleviated the computing power requirements of the Transformers under excellent performance. Our model reduces the trial and error cost of adjusting the segmentation model architecture for specific agricultural applications and provides a potential scheme for high-throughput phenotypic feature detection of plants.

KEYWORDS

cotton leaf segmentation, composite backbone, convolutional neural network, attention mechanism, transformer

1 Introduction

Cotton, the second largest crop after grain, is the primary raw material for daily necessities and the textile industry (Feng et al., 2022). However, biotic stress and abiotic stress existing in cotton production affect the yield and quality (Zhang et al., 2022). To ensure sustainable cotton production, breeders must identify quality varieties through continuous

monitoring of cotton phenotypic traits (Ye, 2014). Budding, flowering, and boll periods are significant growth stages of cotton, which are directly reflected in cotton leaves due to the influence of nutrition, diseases, and insect pests, and thus determine the subsequent growth and yield of cotton (Mubarik et al., 2020). Breeders screen the appropriate cotton varieties during the budding, flowering, and boll period, based on estimates of plant disease resistance and yield reflected by closely related leaf phenotypic traits (e.g., Leaf Length, Leaf Area Index) (Saeed et al., 2021). Manual sampling in complex field environments is a natural way to measure cotton leaf phenotypic parameters. However, manual sampling is a labor-intensive, time-consuming, and disruptive process (Bao et al., 2021). Image segmentation of computer vision is a standard approach for non-destructive sampling samples in complex field environments. The image segmentation algorithm can automatically separate the processed samples to be processed. Therefore, image segmentation has gradually become a potential preprocessing approach of sample separation for rapidly measuring plant phenotypic parameters.

With advances in computing power (e.g., GPU), deep learning with powerful nonlinear and robust generalization ability replaces the traditional image segmentation algorithm, which highly relies on expert experience (Taghanaki et al., 2020). Generally speaking, the segmentation models based on deep learning are composed of encoders and decoders, such as PSPNet (Zhao et al., 2017), DANet (Fu et al., 2019), CPNet (Yu et al., 2020), DeepLab v3+ (Chen et al., 2018). Specifically, the backbones of the segmentation models in the encoder are used to extract features (Miao et al., 2020). The feature diversity of backbone extraction determines the performance of the segmentation model (Minaee et al., 2022). Currently, convolutional neural networks (CNNs, e.g., ResNet-101, Xception) with deep stacked convolution structures to represent powerful features have gradually become mainstream feature extractors. PSPNet utilizes ResNet-101 as a backbone to achieve an elegant expression in the complex field environment of grape segmentation (Chen et al., 2021). DeepLab v3+ employs ResNet-101/Xception as a backbone to segment fruit plaques (Li et al., 2022b; Yuan et al., 2022), and also attempts to segment cotton roots (Kang et al., 2021).

CNNs have been widely used in plant phenotype, especially phenotype segmentation. However, CNNs have apparent disadvantages, such as poor learning ability of low-level features of images and partial neglect of global information, which limit the accurate segmentation of object edges in complex field environments (Liu et al., 2018). Due to the complexity of the leaf environment, the morphological characteristics (texture, size, and shape) of the leaf change accordingly, and the segmentation of the leaf edge has the dilemma of over-segmentation/under-segmentation (Yang et al., 2020). Transformers, as attention models, achieve powerful accuracy for large-scale datasets with a robust representation of global context (Dosovitskiy et al., 2021). In contrast, CNNs with deep stacked convolution structures embedded in the attention modules, e.g., Channel Attention Module (Woo et al., 2018), and Convolution Block Attention Module (Woo et al., 2018), integrate global information to a limited extent, and improve the power slightly of object edge segmentation. Thus, with the success of self-attention models such as Transformers, much previous work has attempted to bring the power of attention to computer vision (Khan et al., 2022).

Recently, Transformer-based backbones have shown potential performance and expanded cutting-edge applications. Li et al. (2022a) proposed an automatic pest recognition method based on Vision Transformer (ViT) in PlantVillage (a public dataset of plant pests and diseases) (Hughes and Salathé, 2015). Reedha et al. (2022) proposed a novel crop recognition model using ViT based on unmanned aerial vehicles (UAV) remote sensing images. Wu et al. (2021) proposed a multi-scale feature extraction model based on a visual converter to identify crop disease types. However, the large model capacity with huge parameters and high computational power required by Transformers hinders rapid application to agricultural tasks (Khan et al., 2022). The attention of Transformers has slight inductive bias and weak generalization on the relatively small amount of datasets compared with the convolution of CNNs (Dosovitskiy et al., 2021).

In relatively small agricultural data sets, plant phenotype researchers have used the Transformer and CNN cascade model, incorporating the inductive bias of CNNs and the self-attention mechanism of Transformers, to study plant phenotype. Wang et al. (2022) proposed a crop segmentation method of remote sensing images based on a barely remote sensing dataset by constructing a novel architecture of coupling CNN and Transformer. Liu et al. (2022) attempted to propose a CNN-Transformer network with Multi-Scale Context Aggregation (MSCANet) and realize efficient and effective farmland change detection. However, Transformer and CNN cascade models integrate the respective advantages of Transformers and CNNs, and the computational cost and data requirements of Transformers are also introduced into the cascade models, which hinders the rapid promotion of the cascade models in agriculture. Therefore, for the global learning potential of the self-attention mechanism of Transformers and the fast application limitation of Transformers required computing power and large-scale datasets, the models combining convolution of CNNs and self-attention of Transformers have become a new research direction. CoAtNet (Dai et al., 2021), as a novel backbone, incorporates the global awareness of Transformers and the inductive bias of CNNs.

Different from Transformer and CNN Cascade Models, CoAtNet introduces CNN convolution and Transformer attention to alleviate computational power greed. The classification speed and accuracy of CoAtNet in ImageNet demonstrate the potential of CoAtNet as a backbone for segmentation models. However, the robust backbone design of the segmentation models requires substantial trial-and-error costs (e.g., expert experience and computational costs). As the backbone architecture of automatic search, neural architecture search (NAS) (Zoph and Le, 2017) still has the computational cost of architecture search. Therefore, for backbone design, simple and effective strategies are urgently needed for rapid application in agriculture. CBNet (Liu et al., 2020) and CBNetV2 (Liang et al., 2021) proposed the architectures integrating multiple backbones into a composite backbone for object detection, which assembles multiple existing backbones in parallel to represent various features, reducing the computational cost of architecture design. Inspired by CBNet and CBNetV2, a leaf segmentation architecture based on composite backbone architecture was proposed and explored.

To the best of our knowledge, the encoder-decoder architecture segmentation model has over-segmentation and under-segmentation

in complex field environments. Among the encoders of the segmentation models, the design of a robust backbone can alleviate segmentation anomalies, especially the mainstream CNNs and Transformers. CNNs are highlighted by inductive learning and generalization, while Transformers are highlighted by global semantics. However, Transformers and cotton-leaf segmentation architecture design is power consumption. Therefore, this work aims to explore the application of the composite backbone architecture combined with the convolution of CNNs and the attention of Transformers in cotton leaf segmentation without significantly introducing the computational power requirements of Transformers. The specific objectives achieved herein are as follows:

(1) Eight hundred images of budding, flowering, and boll period cotton leaves in five typical complex field environments (normal, spotted lesions, regional lesions, occluded blades, uneven illumination) were collected and labeled.

(2) CoAtNet, which incorporates the attention mechanism of Transformers into the convolution, was explored as the backbone of the encoder in the cotton leaf segmentation architecture.

(3) A simple and effective composite backbone (Xception and CoAtNet) leaf segmentation architecture combining convolution and attention was designed to fully learn the edge information and global context of cotton leaves.

An outline is employed to show the detailed steps of this work in Figure 1. Our model is based on the encoder-decoder architecture of DeepLab v3+, and the composite backbone is introduced into our model. In step 1, Xception and CoAtNet are used as the lead backbone and assisting backbone in the composite backbone, and the features of the input image are first extracted by assisting backbone. In step 2, the output features of each stage of the assisting backbone flow to parallel and lower stages of the lead backbone. Xception learns the richer multi-level features of the assisting backbone. In step 3, the fusion mechanism of weight contribution factors is adopted to suppress unimportant features from different backbones. The fused features flow to the lead backbone under the batch-normalized channel weight contribution factor. Finally, the output of the composite backbone is applied to the encoder and decoder.

2 Materials and methods

In this section, Section 2.1 introduced the subdivision of the cotton dataset into acquisition and preprocessing. Then, Section 2.2 illustrated the design of the segmentation model, including the model framework and the composite backbone. Finally, Section 2.3 introduced the experimental details, including the experimental structure, training, and testing strategy.

2.1 Data description

2.1.1 Acquisition

Cotton crops were grown in the field at the experimental station (85°9′51.231 00′′E, 44°35′47.720 00′′N) of the Agricultural College of Shihezi University, Shihezi, China. The cotton variety “Xinluzao 54” was trial-planted on April 7, 2021, and the sowing density was ten seeds/square meter. Specifically, the column spacing was 0.2 m, and the row spacing was 0.3 m. The images were acquired along the rows over the entire field on six experimental dates in the budding, flowering and boll period (June 11, June 18, June 23, July 7, July 13, and July 22). Multiple smartphones were selected to capture images and verify the generality of the subsequent segmentation models. The smartphone cameras were set to manual operation mode, with a distance of about 0.3 m from the target leaves. Specifically, the target leaves were photographed in natural light (9:00-12:00 a.m., Beijing Time). The following five types of cotton leaves were typical research objects, as shown in Figure 2.

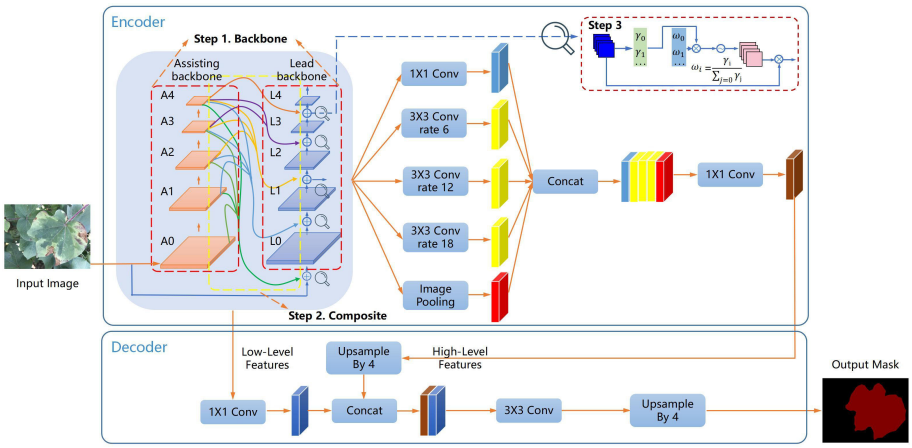


FIGURE 1
The outline of the composite backbone and segmentation architecture in our model. In step 1, the composite backbone (Xception and CoAtNet) is selected, and the features of the input image are extracted by CoAtNet. In step 2, multi-scale features are interacted in the composite backbone. In step 3, features from the composite backbone are fused using weight contribution factors.

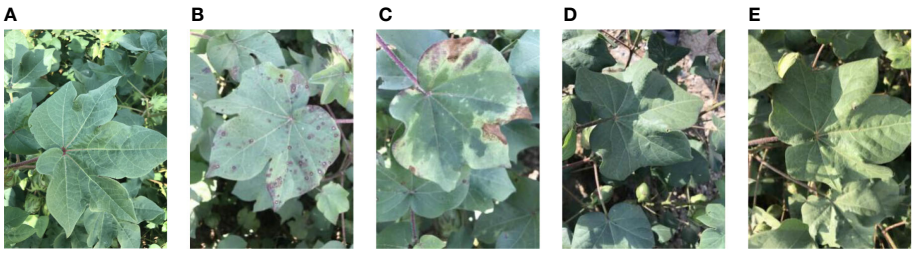


FIGURE 2
Images of the Cotton Leaf dataset. The dataset is divided into five representative leaves: **(A)** a normal cotton leaf, **(B)** a cotton leaf with spotted lesions, **(C)** a cotton leaf with regional lesions, **(D)** a cotton leaf with occluded blades, and **(E)** a cotton leaf with uneven illumination.

- Normal leaves;
- Leaves with spotted lesions;
- Leaves with regional lesions;
- Leaves with occluded blades;
- Leaves with uneven illumination.

2.1.2 Preprocess

The median filtering algorithm was applied to image preprocessing since a certain amount of image noise caused by external factors would negatively impact the training of segmentation models. Moreover, the image resolution was adjusted to 512×512 pixels before annotation, saving computational resources and labor handling time. Subsequently, the polygons pattern in Labelme-3.3.6 (Torralba et al., 2010) provided labels for two semantic classes of the dataset, including foreground (target leaves) and background (i.e., soil, weeds, other leaves). The image annotation process is shown in Figure 3. The diversity of leaf images under different growth periods was considered, and at least 100 images were labeled from five typical cotton leaves in the budding, flowering, and boll period.

The size and diversity of the dataset affect the segmentation model performance (Barbedo, 2018). Specifically, large-scale datasets are a prerequisite for building reliable segmentation models, while limited datasets easily lead to model overfitting. Therefore, a series of operations was adopted to expand the cotton leaf dataset: rotation and mirror flip. The final cotton leaf dataset containing 800 images and segmentation labels was divided into 80% training dataset and 20% testing dataset for training and testing subsequent segmentation models.

2.2 Model design

2.2.1 Framework

Currently, the performing segmentation models rely heavily on the backbones. Intuitively, the rich feature maps extracted by the backbones and the vast receptive fields sensed by the backbones determine the segmentation model performance (Ma et al., 2020). However, designing and pre-training a new backbone consumes various computing resources, and requires a large number of training samples (Bao et al., 2022). Recently, the application of composite backbone in object detection has inspired our model (Liang et al., 2021). A composite backbone combines several existing networks and then integrates the rich features of multiple scales. In addition, previous studies have shown that the feature pyramid network (FPN) is more effective than simple network deepening or broadening. Top-down paths of FPN introduce spatially richer and semantically more powerful high-level features and enhance low-level features in bottom-up paths of FPN. Thus, in our model, multiple backbones are composited and called assisting backbone and lead backbone, respectively. The composite backbone of our model extended FPN (Lin et al., 2017) idea combines high-level and low-level features from multiple networks.

As a classical semantic segmentation model, DeepLab v3+ (Chen et al., 2018) is used as the benchmark for segmentation models. Therefore, DeepLab v3+ is regarded as the prototype of our model, and the lead backbone is the Xception applicable to segmentation in the raw DeepLab v3+. However, DeepLab v3+ still does not fully show excellent potential performance and only tries mature convolutional neural networks (CNN) as a backbone. As a Backbone, simple CNN has the problems of missing global information and tiny local receptive fields, which cannot meet the requirements of DeepLab v3+ for feature maps. In addition, CoAtNet (Dai et al., 2021) integrates the attention mechanism of Transformers into the convolution operation of CNN, maintaining the optimal tradeoff between model generalization capability and model capacity. Therefore, the hybrid family of CoAtNet is used as the assisting backbone of our model (based on DeepLab v3+).

As shown in Figure 1, our model is based on the encoder-decoder architecture of DeepLab v3+. Our model uses Xception (Chen et al., 2018) and CoAtNet (Dai et al., 2021) as the lead backbone and assisting backbone. In addition, our model is inspired by FPN and contains long-skip connections from the encoding path to the decoding path and short-skip connections between the composite backbone. Long-skip connections transmit low-level features and high-level features. Short-skip connections fuse assisting backbone and lead backbone features, and transmit to the lead backbone.

The remaining parts retain the original architecture of DeepLab v3+. The encoder of the atrous spatial pyramid pool (Chen et al., 2017) module processes the lead backbone output features with five different operations, namely 1×1 convolution, 3×3 convolution at dilation rate 6, 3×3 convolution at dilation rate 12, 3×3 convolution at dilation rate 18, and Image Pooling. The output features of five different operations are downsampled to 1/16 of the input image size and then combined to form multi-scale features. The multi-scale features are then subjected to 1×1 convolution operation to form high-level features. The low-level features output by the assisting backbone A1 are combined and fused with the high-level features four



FIGURE 3
Image annotation process. The left is the input image, and the right is the labeled image.

times up-sampled after the 1×1 convolution operation. The low and high-level fusion features are restored to the input image size by 3×3 convolution and four times upsampling. In our model, two dropout layers are added before the last four times upsampling layers to avoid overfitting. The softmax function finally activates our model. Each channel value of the activation output represents the category probability, and the maximum probability value determines the pixel category.

2.2.2 Backbone

Our model is based on CoAtNet and Xception as the composite backbone. As shown in Figure 4, the official Xception backbone for segmentation is retained as the lead backbone. In our model, the lead backbone and assisting backbone are divided into five standard blocks, which are L0, L1, L2, L3, and L4 of the lead backbone, and A0, A1, A2, A3, and A4 of the assisting backbone in turn. Concretely, our model divides Xception into five modules, L0, L1, L2, L3, and L4, according to the remaining residual connection after the first residual connection. Modules L0, L1, L2, L3, and L4 are composed of only 3×3 separable convolution to reduce computational power requirements. The L3 module is repeated 16 times to learn the image features fully. The rest consists of 3×3 convolution and 3×3 separable convolution. 1×1 convolution achieves feature channel rise and residual transfer. In Xception, the number of channels of the feature map increases successively, and the partial convolution step is set to 2 to fully capture the spatial information of the feature map and reduce the spatial resolution.

As shown in Figure 5, in our model, the assisting backbone consists of three convolution modules, A0, A1, and A2, and two self-attention modules, A3 and A4. The A0 module consists only of 3×3 convolution, which reduces the feature spatial resolution. Modules A1 and A2 are expanded by the attention mechanism of MobileNet consisting of 1×1 convolution and 3×3 separable convolution (MBConv module with inverted bottleneck structure) (Sandler et al., 2018). 1×1 convolution is used to increase and reduce the dimension of the feature. A3 and A4 modules contain a Relative-Attention (Rel-Attention) layer and a Feed-Forward Network (FFN) layer for learning global feature information. The modules A1, A2, A3, and A4, are successively repeated 2, 4, 8, and 2 times to explore the features fully. The rest consists of global pooling and a fully connected (FC) layer. The residual connection is guaranteed to reduce the model complexity to reduce overfitting, while the residual connection prevents the gradient from disappearing. Specifically, 1×1 convolution carries out feature channel dimension raising and completes the residual transfer.

The Rel-Attention layer expands the attention mechanism of Transformers. The Rel-Attention layer stretches the input features from three-dimensional to two-dimensional, that is, h×w×c to (h×w)×c, and then gets the Input Embeddings. The trainable weight matrices of Queries, Keys, and Values are calculated by the

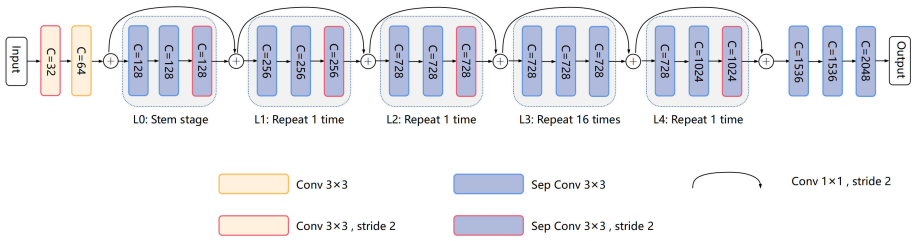


FIGURE 4
Xception. C represents the number of feature output channels.

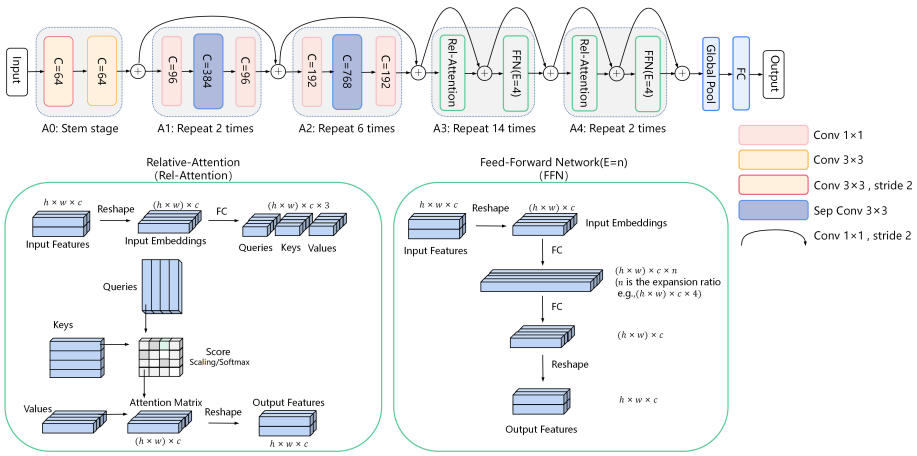


FIGURE 5
CoAtNet. CoAtNet is divided into five modules: three convolution modules, A0, A1, A2, and two self-attention modules, A3 and A4. C represents the number of feature output channels. E is the n-time expansion rate of the Feed-Forward Network (FFN) layer.

Input Embeddings with the full connection. Intuitively, the two-dimensional matrix Queries, Keys, and Values all contain feature global information. The Score matrix is computed by the scalar product of Queries and Keys. The Score matrix represents the correlation between each one-dimensional vector in Keys and each one-dimensional vector in Queries. Further, the Score matrix is scaled and activated by the softmax function. Then, the Attention Matrix is obtained by calculating the scalar product between the Score matrix and Values, which contain relative global attention features of each one-dimensional vector in the three matrices of Queries, Keys, and Values. Finally, the Attention Matrix is reconverted into three dimensions to obtain the output features.

The FFN layer learns advanced image features from the MBConv block. The Input Embeddings are expanded by an FFN layer consisting of multiple FC layers with an n- time expansion rate and then resized to the original size. In our model, the number of feature channels in the FFN layer inflation factor was set to 4.

2.2.3 Composite

Backbone, or feature extractor, as the initial stage of the semantic segmentation network, plays a significant role in model segmentation performance (Fan et al., 2018). Backbone provides the basic features of the segmentation target for the semantic segmentation model. Our model draws on the ideas of FPN (Lin et al., 2017) and CBNetV2 (Liang et al., 2021) architecture to construct the connection structure between the lead backbone and the assisting backbone. As shown in Figure 6, the output features of modules A0, A1, A2, A3, and A4 of CoAtNet flow to parallel and lower-level jump connections of Xception. Xception both preserves the original residual connection and learns the richer multi-level features of the assisting backbone. Specifically, the output feature maps of modules A0, A1, A2, A3, and A4 are consistent with the dimension of the output feature maps of Xception and skip-connections of lower stages by 1×1 convolution. Subsequently, linear interpolation keeps the output feature maps of A0, A1, A2, A3, and A4 modules consistent with the spatial resolution of the output feature maps at parallel and lower skip-connections of Xception. Finally, the output feature maps of modules A0, A1, A2, A3, and A4 are element-summed with the output feature maps at parallel and lower-level skip-connections of Xception.

The output of each stage of the assisting backbone flows to parallel and lower stages of the lead backbone. The output of the lead backbone is applied to downstream tasks. Different from the simple network deepening or broadening, the composite backbone, which integrates the high and low-level features of the composite backbone, gradually expands the receiving field and provides richer target information. Due

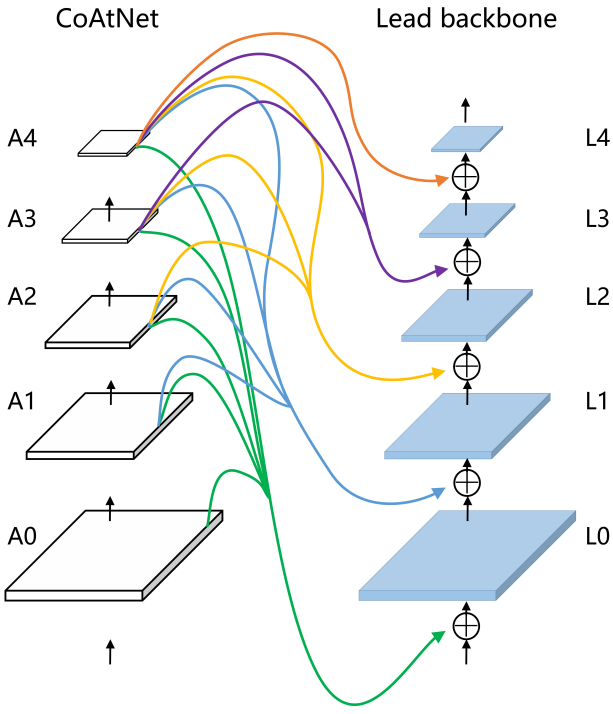


FIGURE 6
Our composite backbone architecture with CoAtNet as assisting backbone.

to the different response values of the multi-level features integrating the composite backbone, the model is prone to convergence dilemmas. Inspired by the accelerated convergence of normalization (Yan et al., 2020), our model adopts the fusion mechanism of weight contribution factors to suppress unimportant features, as shown in Figure 7. The fused features flow to the lead backbone of Xception under the batch-normalized channel weight contribution factor.

2.3 Experiment

2.3.1 Experimental detail

2.3.1.1 Hardware

Experiments were conducted with the following hardware configurations: Intel(R) Core(TM) i7-11700 K CPU, 128GB memory, and NVIDIA GeForce RTX3090 graphics card.

2.3.1.2 Software

The deep learning framework PyTorch installed in Windows 10 (Microsoft, United States) was adopted to build neural network models.

2.3.1.3 Loss function

Models were optimized by the cross-entropy loss (cost) function (Huang et al., 2016). As shown in Equation (1), y_i represents the label of the pixel, p_i represents the predicted value of the pixel, and m represents the number of pixels in the image.

$$Loss = \sum_{i=1}^{m} -(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (1)$$

The composite backbone was applied in our model to train the original cross-entropy loss. The assisting backbone, which inherited the assistant loss concept of CPNet, was also used to produce assistant supervision. In other words, original cross-entropy loss bears the greatest responsibility, and assistant supervision helps to optimize the learning process. Meanwhile, super parameter weight was added to balance the assistant supervision. The loss defined in our model is as Equation (2).

$$L = L_{Comp} + \lambda \cdot L_{Assist} \quad (2)$$

Where L_{Comp} is the loss of the composite backbone from input to output,is the loss of assisting backbone from the input only through the low-feature path to the output, and λ is the super parameter weight for the assistant supervision. In our model, λ was set to 0.3 according to our empirical experiments.

2.3.2 Training strategy

Two training strategies were used on the cotton leaf dataset for our model. In the first strategy, our model was trained from scratch. In the second strategy, to use the leaf information of the source domain and effectively transfer knowledge to the target domain, the PlantVillage (Hughes and Salathé, 2015) dataset consisting of crop leaf images was first used to pre-train the lead backbone and the assisting backbone. The composite backbone with pre-trained weights in a fine-tuning paradigm of the training process to achieve fast learning on the cotton leaf dataset. In particular, in the fine-tuning paradigm, the composite backbones were frozen to train the encoder-decoder part of our model fully. Then, the composite backbones were unfrozen to complete the rest after the model was trained for a certain epoch.

The parameter setting in training from scratch is shown in Table 1, and the parameter setting in fine-tuning is shown in Table 2. The optimizer of our model was the adaptive moment estimation optimizer (Adam) (Kingma and Ba, 2015). In Adam, the first and second moments of the gradient were used to update and correct the current learning rate (Dong et al., 2017). More importantly, if the loss did not improve for more than five epochs during the training, the minimum learning rate was set to 0. Otherwise, the learning rate would drop by 1/2, and the model would continue to train at that learning rate. The model would stop training until the loss no longer changes significantly or until the maximum number of iterations was reached.

2.3.3 Testing strategy

Pixel Accuracy (PA), Mean Pixel Accuracy (MPA), and Mean Intersection over Union (MIoU) (Shelhamer et al., 2015) are used to evaluate the effect of our model, as shown in Equation (3), (4) and (5).

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \quad (3)$$

$$MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \quad (4)$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \quad (5)$$

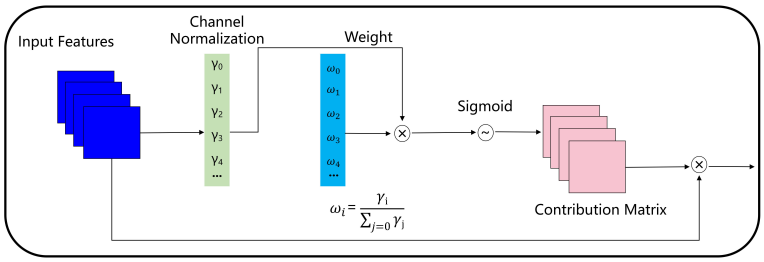


FIGURE 7
Fusion mechanism of weight contribution factors based on batch normalization. Where, γ_i represents the weight value of the i-th channel calculated in batch normalization, γ_j represents the weight value of the j-th channel calculated in batch normalization, ω_i represents the importance degree of the i-th channel.

TABLE 1 The parameter setting in training from scratch.

Optimizer	Learning rate	Batch size	Epochs
Adam	5e-4	4	200

TABLE 2 The parameter setting in fine-tuning training.

Training Stage	Optimizer	Learning rate	Batch size	Epochs
Backbone freezing	Adam	1e-4	8	100
Fine-tuning	Adam	5e-5	4	100

Where, k represents the number of classes, i represents the true value, j represents the predicted value, and p_{ij} represents the pixels that predict class i as class j . Generally, p_{ii} represents real samples (TP), p_{ij} represents false negative samples (FN), and p_{ji} represents false-positive samples (FP).

However, the MIoU score is higher than the true value when measuring the boundary quality, which cannot gracefully evaluate the segmentation results of our model. Accordingly, Boundary Intersection over Union (BIoU) is introduced as an additional evaluation metric to compare the segmentation fineness better (Cheng et al., 2021). BIoU is used to evaluate the boundary quality of segmented objects based on the sensitivity of boundary error. BIoU is defined as Equation (6).

$$BIoU = \frac{|(G_d \cap G) \cap (P_d \cap P)|}{|(G_d \cap G) \cup (P_d \cap P)|} \quad (6)$$

Where G denotes the ground truth binary mask, P denotes the prediction binary mask, and d denotes the pixel width of the boundary region. Boundary regions G_d and P_d are the sets of all pixels within d pixels distance from the ground truth and prediction contours, respectively.

3 Results and discussion

3.1 Segmentation model comparison experiment

Segmentation models adopt the experimental setting in Section 2.3.2 for training to make the comparison fair. The performance of segmentation models in training from scratch is shown in Table 3, and the implementation of segmentation models in fine-tuning is shown in Table 4.

Compared with the encouragement success of training from scratch, the evaluation indexes (BIoU and MIoU) of each segmentation model in fine-tuning training were improved accordingly. In addition, among the two training strategies, PSPNet fused multi-scale features to obtain the baseline effect in the cotton leaf segmentation task under complex background. DANet inherited the attention mechanism to improve the cotton leaf segmentation task. CPNet had achieved moderate results without multi-scale feature fusion and attention mechanism, considering assistant supervision strategy. DeepLab v3+ took a mature CNN (Xception) as a backbone, which was the benchmark level in several standard segmentation models, both in MIoU, which represented the overall segmentation quality of the cotton leaf, and in BIoU, which meant the segmentation quality of the leaf edge.

Our model had significant progress compared with DeepLab v3+. Specifically, among MIoU with already high ratings, our model increased by about 1%, due to data limitations or task bottlenecks with an inconspicuous rise. However, in BIoU, our model improvement was quite noticeable, with an increase of around 5%. Without loss of generality, the BIoU was enhanced due to the composite backbone (Xception + CoAtNet). The introduction of our composite backbone not only guaranteed the generalization ability and convergence ability based on Xception, but also had the global receptive field of the self-attention layer based on CoAtNet. The global information ensured that our model worked more accurately in cotton leaf edge segmentation. Due to the structure of the composite backbone, multi-level features were obtained by the encoder and decoder of our model, thus enabling the edge pixel predictor to get a rich feature map. In addition, our model considered the progress of CPNet, which also increased the weight of our assisting loss. At the same time, the composite backbone architecture retained the conventional training mode of the

TABLE 3 The performance of segmentation Models in training from scratch.

Method	Backbone	Multi-scale Fusion	Attention	Assistant Supervision	BIoU	MIoU	MPA	PA
PSPNet	ResNet-101	○	×	×	0.415	0.826	0.869	0.877
DANet	ResNet-101	×	○	×	0.488	0.883	0.917	0.933
CPNet	ResNet-101	×	×	○	0.497	0.896	0.927	0.941
DeepLabv3+	Xception	○	×	×	0.522	0.911	0.951	0.967
Ours	Composite (Xception + CoAtNet)	○	○	○	**0.583**	**0.924**	**0.964**	**0.972**

The bold values indicate the maximum value in their columns.

TABLE 4 The performance of segmentation Models in fine-tuning training.

Method	Backbone	Multi-scale Fusion	Attention	Assistant Supervision	BIoU	MIoU	MPA	PA
PSPNet	ResNet-101	◯	×	×	0.438	0.866	0.893	0.901
DANet	ResNet-101	×	◯	×	0.513	0.899	0.925	0.943
CPNet	ResNet-101	×	×	◯	0.533	0.911	0.937	0.953
DeepLabv3+	Xception	◯	×	×	0.565	0.923	0.957	0.972
Ours	Composite (Xception + CoAtNet)	◯	◯	◯	**0.608**	**0.940**	**0.975**	**0.979**

The bold values indicate the maximum value in their columns.

backbone in essence. Decoupling the composite backbone and then pre-training the weight of the individual backbones independently was low-cost.

3.2 Segmentation model robust experiment

To make the comparison concrete, various images from the test set of the cotton leaf dataset were selected to visualize the results of the pre-trained segmentation models, and the types of cotton leaf images were described in Section 2.1.1. The comparison results are shown in Figure 8. The segmentation models effectively detect normal and diseased cotton leaves (spotted and regional lesions), especially in detecting cotton leaf edges. The texture features and shape parameters of the cotton leaves during training were simple to learn. Under the condition of shadow occlusion, the overall segmentation of our model and DeepLab v3+ was satisfactory. At the same time, CPNet had the under-segmentation phenomenon, DANet and PSPNet had the over-segmentation and under-segmentation phenomenon. DeepLab v3+, CPNet, DANet, and PSPNet over-segmented cotton leaves compared with the segmentation acceptable to our model under uneven illumination conditions.

The PSPNet, with ResNet-101 as the backbone, incorporated multi-scale features. The segmentation of normal and diseased cotton leaves (spotted and regional lesions) was consistent with the further determination of cotton phenotypic traits. DANet integrated with the attention mechanism, similar to PSPNet, and both had under-segmentation under the condition of shadow occlusion and uneven illumination. CPNet and DeepLab v3+, in turn, due to the backbone update and the introduction of assistant losses, the overall segmentation level was moderately acceptable except for under-segmentation in shadows and over-segmentation in uneven illumination. Since the conventional segmentation models only contained the convolution module and lacked the global receptive field, the conventional segmentation models could not learn the subtle differences between pixels. The processing effects of leaf edges were poor in the complex filed environment.

In contrast, our model based on DeepLab v3+ accurately segmented cotton leaves in typical scenes, especially the edge of cotton leaves. Due to the proper coordination of convolution and self-attention module of assisting backbone CoAtNet and the penalty of assisting loss, our composite model could effectively learn the local and global context of complex background. The excellent performance of our model cannot be achieved without the self-attention module in the assisting backbone. In addition, our model inherited the idea of the various benchmark models to ensure that the encoder had full access to the information from the multi-layer features.

3.3 Ablation experiment

The assistant supervision in our model ensured that the assisting backbone contributed to the segmentation. Therefore, the penalty of assisting loss enables the model to learn more cotton leaf features, as CPNet achieved satisfactory improvement by only considering assistant loss. In addition, to fairly compare the progress of our model with DeepLab v3+, the results of decoupled assistant supervision are shown in Figure 9. Figure 9 shows the improvement effect of assistant supervision in training from scratch and fine-tuning training strategies. In the training-from-scratch strategy, MIoU and BIoU improved from 0.915 to 0.924, and 0.553 to 0.583, respectively. Accordingly, in the fine-tuning training strategy, MIoU and BIoU improved from 0.929, and 0.585 to 0.940 and 0.608, respectively.

In training from scratch and fine-tuning training strategies, the trajectory occasionally shows sudden declines. One of the reasons for the decline phenomenon may be the random loading of batch samples in the training data set to train our model. The randomness of training samples led to significant fluctuations in the parameters of our model, which further affected the performance of our model on the test dataset. Besides, to prevent the model from overfitting, two dropout layers were added before the upsampling layer of the decoder. Although the dropout layers can improve the robustness of our model, the dropout layers cause important neurons to be randomly deactivated, which would be the reason for the sudden declines of the trajectory. However, the introduction of assisted supervision promoted the segmentation power of our model, and the training was smoother than that of the non-assisted supervision strategy. The trajectory can recover and rise in fewer epochs after a sudden decline with assisted supervision. The segmentation effect of our model was suboptimal without adopting the assisted supervision strategy. Generally, the attention mechanism of Transformers integrated into the composite backbone of our model achieved remarkable results. Due to the limitation of computing resources, the computational requirements of the Transformer cannot be met.

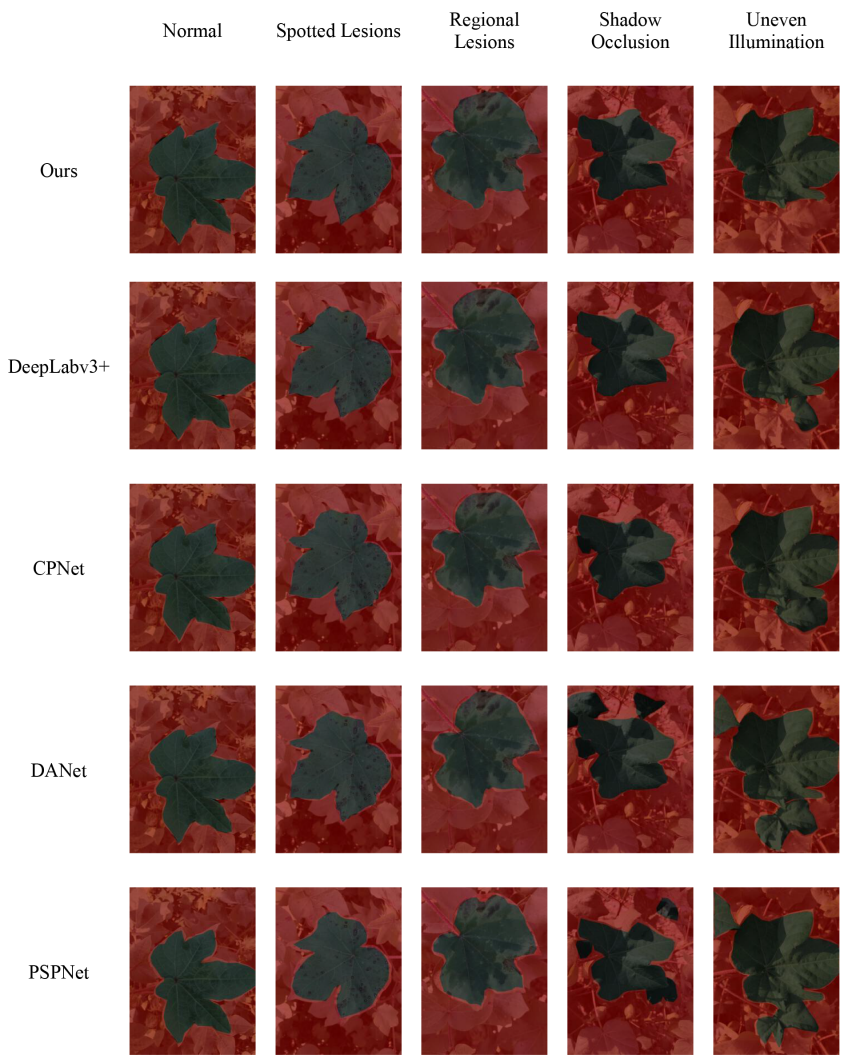


FIGURE 8
Pre-trained segmentation models results on five types of cotton leaf images.

Further, in the ablation experiment, the assisting backbone is replaced by the Transformer for comparison with our model. However, our model incorporated the attention mechanism for the broad success of Transformers, which provides a feasible strategy for overcoming the computational power requirements of Transformers and applying Transformers elegantly to agricultural tasks.

4 Conclusion

In this work, from five typical cotton leaves (normal, spotted lesions, regional lesions, occluded blades, uneven illumination), a total of 800 images were labeled at the budding, flowering, and bolling stages. The composite backbone-based encoder and decoder semantic segmentation architecture (our model) was used for cotton leaf segmentation in complex field environments. The composite backbone consisted of the lead backbone Xception and the assisting backbone CoAtNet, saving the computational cost of architecture search for cotton-leaf segmentation. Xception represented the biased learning and generalization of CNN, CoAtNet was integrated into our model with the global context inherited from Transformers. Due to the slight computational power and data requirements of CoAtNet compared with Transformers, our model not only maintained the fast convergence of convolution but also maintained the global receptive field of attention under the constraint of a certain computational cost. At the same time, the introduction of the multi-scale feature fusion mechanism and assistant supervision strategy effectively improved the performance of our model. The experimental results showed that the cotton leaf segmentation performance of our model, especially under complex filed environments, was significantly better than that of the PSPNet, DANet, CPNet and DeepLab v3+ benchmark models, and the under-segmentation and over-segmentation of five typical cotton leaves were encouraging. In addition, different backbones can be trained offline and reassembled into composite backbones with limited computing resources. In the future, more types and numbers of pre-trained backbones can be combined to achieve faster and better plant high-throughput phenotypic tasks.

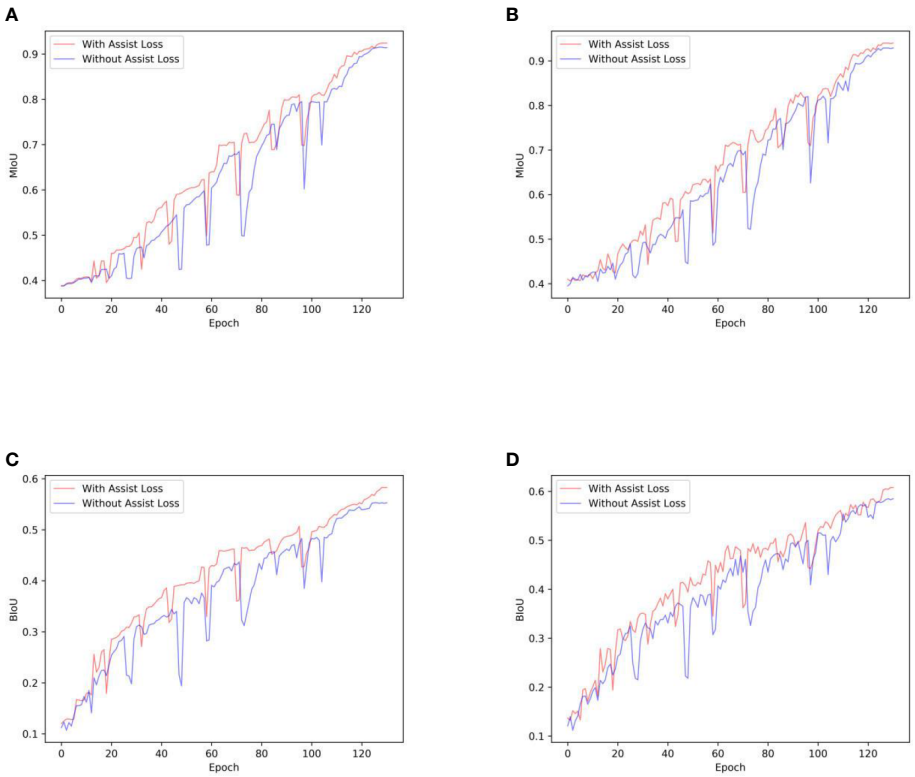


FIGURE 9
The effects of assistant supervision: **(A, C)** represent the changing trends of MIoU and BIoU in training from scratch, **(B, D)** are the changing trends of MIoU and BIoU in fine-tuning training.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JY, TY, PG, and WX contributed to conception and design of the study. JY, TY, and WY contributed to the preparation of equipment and the acquisition of data. JY and TY wrote the code and tested the method. JY, TY, WY, PG, and WX validated the results. JY wrote the first draft of the manuscript. TY, PG, WX, and XL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work has been partially supported by the National Natural Science Foundation of China (grant numbers, 61965014 and 62265015), and the Postgraduate Scientific Research and Innovation Project of Xinjiang Uygur Autonomous Region (grant number, XJ2022G107).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bao, W., Cui, Q., Chen, B., and Yang, B. (2022). Phage_UniR_LGBM: Phage virion proteins classification with UniRep features and LightGBM model. *Comput. Math. Methods Med.* 2022. doi: 10.1155/2022/9470683

Bao, W., Yang, B., and Chen, B. (2021). 2-hydr_Ensemble: Lysine 2-hydroxyisobutyrylation identification with ensemble method. *Chemometrics Intelligent Lab. Systems.* 215, 104351. doi: 10.1016/j.chemolab.2021.104351

Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Engineering.* 172, 84–91. doi: 10.1016/J.BIOSYSTEMSENG.2018.05.013

Cheng, B., Girshick, R. B., Doll'ar, P., Berg, A. C., and Kirillov, A. (2021). "Boundary IoU: Improving object-centric image segmentation evaluation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: Computer Vision Foundation / IEEE). 15329–15337. doi: 10.1109/CVPR46437.2021.01508

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. intelligence.* 40, 834–848. doi: 10.1109/TPAMI.2017.2699184

Chen, S., Song, Y., Su, J., Fang, Y., Shen, L., Mi, Z., et al. (2021). Segmentation of field grape bunches *via* an improved pyramid scene parsing network. *Int. J. Agric. Biol. Engineering.* 14 (6), 185–194. doi: 10.25165/j.ijabe.20211406.6903

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Eur. Conf. Comput. Vision* (Berlin: Springer), 11211:801–818. doi: 10.1007/978-3-030-01234-2_49

Dai, Z., Liu, H., Le, Q. V., and Tan, M. (2021). CoAtNet: Marrying convolution and attention for all data sizes. *Adv. Neural Inf. Process. Syst.* 34, 3965–3977. doi: 10.48550/arXiv.2106.04803

Dong, H., Yang, G., Liu, F., Mo, Y., and Guo, Y. (2017). "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in *annual conference on medical image understanding and analysis*, (Berlin: Springer). 506–517. doi: 10.1007/978-3-319-60964-5_44

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, (Ithaca, NY: OpenReview.net.). doi: 10.48550/arXiv.2010.11929

Fan, L., Wang, W.-C., Zha, F., and Yan, J. (2018). Exploring new backbone and attention module for semantic segmentation in Street scenes. *IEEE Access.* 6, 71566–71580. doi: 10.1109/ACCESS.2018.2880877

Feng, L., Chi, B., and Dong, H.-Z. (2022). Cotton cultivation technology with Chinese characteristics has driven the 70-year development of cotton production in China. *J. Integr. Agriculture.* 21 (3), 597–609. doi: 10.1016/S2095-3119(20)63457-8

Fu, J., Liu, J., Tian, H., Fang, Z., and Lu, H. (2019). "Dual attention network for scene segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: Computer Vision Foundation / IEEE). 3141–3149. doi: 10.1109/CVPR.2019.00326

Huang, C., Li, Y., Loy, C. C., and Tang, X. (2016). "Learning deep representation for imbalanced classification," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Los Alamitos, CA: IEEE Computer Society). 5375–5384. doi: 10.1109/CVPR.2016.580

Hughes, D., and Salathé, M. (2015). "An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing," in *arXiv*. Available at: https://arxiv.org/abs/1511.08060.

Kang, J., Liu, L., Zhang, F., Shen, C., Wang, N., and Shao, L. (2021). Semantic segmentation model of cotton roots in-situ image based on attention mechanism. *Comput. Electron. Agric.* 189, 106370. doi: 10.1016/j.compag.2021.106370

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision: A survey. *ACM computing surveys (CSUR).* 54, 1–41. doi: 10.1145/3505244

Kingma, D. P., and Ba, J. (2015). "Adam: A method for stochastic optimization," in *Anon. InternationalConferenceon Learning Representations* 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA (Ithaca, NY: OpenReview.net). doi: 10.48550/arXiv.1412.6980

Liang, T., Chu, X., Liu, Y., Wang, Y., Tang, Z., Chu, W., et al. (2021). "CBNetV2: A composite backbone network architecture for object detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: Computer Vision Foundation / IEEE). doi: 10.48550/arXiv.2107.00420

Li, H., Li, S., Yu, J., Han, Y., and Dong, A. (2022a). "Plant disease and insect pest identification based on vision transformer," in *International Conference on Internet of Things and Machine Learning (IoTML 2021)*, (New York, NY: ACM). 194–201. doi: 10.1117/12.2628467

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, (Los Alamitos,CA: IEEE Computer Society). 2117–2125. doi: 10.1109/CVPR.2017.106

Liu, M., Chai, Z., Deng, H., and Liu, R. (2022). A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection. *IEEE J. Selected Topics Appl. Earth Observations Remote Sensing.* 15, 4297–4306. doi: 10.1109/jstars.2022.3177235

Liu, R., Lehman, J., Molino, P., Such, F. P., Frank, E., Sergeev, A., et al. (2018). An intriguing failing of convolutional neural networks and the CoordConv solution. *Adv. Neural Inf. Process. systems.* 2018, 31. doi: 10.48550/arXiv.1807.03247

Liu, Y., Wang, Y., Wang, S., Liang, T., Zhao, Q., Tang, Z., et al. (2020). "Cbnet: A novel composite backbone network architecture for object detection," in *Proceedings of the AAAI conference on artificial intelligence*, (Menlo Park: AAAI Press). 11653–11660. doi: 10.1609/aaai.v34i07.6834

Li, K., Zhang, L., Li, B., Li, S., and Ma, J. (2022b). Attention-optimized DeepLab V3 + for automatic estimation of cucumber disease severity. *Plant Methods* 18, 1–16. doi: 10.1186/s13007-022-00941-8

Ma, W., Cao, Y., Bao, W., Yang, B., and Chen, Y. (2020). ACT-SVM: Prediction of protein-protein interactions based on support vector basis model. *Sci. Programming.* 2020, 1–8. doi: 10.1155/2020/8866557

Miao, C., Xu, Z., Rodene, E., Yang, J., and Schnable, J. C. (2020). Semantic segmentation of sorghum using hyperspectral data identifies genetic associations. *Plant Phenomics* 2020. doi: 10.34133/2020/4216373

Minaee, S., Boykov, Y., Porikli, F. M., Plaza, A. J., Kehtarnavaz, N., and Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intelligence.* 44, 3523–3542. doi: 10.1109/TPAMI.2021.3059968

Mubarik, M. S., Ma, C., Majeed, S., Du, X., and Azhar, M. T. (2020). Revamping of cotton breeding programs for efficient use of genetic resources under changing climate. *Agronomy.* 10, 1190. doi: 10.3390/agronomy10081190

Reedha, R., Dericquebourg, E., Canals, R., and Hafiane, A. (2022). Transformer neural network for weed and crop classification of high resolution UAV images. *Remote. Sens.* 14, 592. doi: 10.3390/rs14030592

Saeed, F., Khan, M. A., Sharif, M., Mittal, M., Goyal, L. M., and Roy, S. (2021). Deep neural network features fusion and selection based on PLS regression with an application for crops diseases classification. *Appl. Soft Comput.* 103, 107164. doi: 10.1016/j.asoc.2021.107164

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Piscataway, NJ: Computer Vision Foundation / IEEE). 4510–4520. doi: 10.34133/2022/9787643

Shelhamer, E., Long, J., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Los Alamitos,CA: IEEE Computer Society). 3431–3440. doi: 10.1109/CVPR.2015.7298965

Taghanaki, S. A., Abhishek, K., Cohen, J. P., Cohen-Adad, J., and Hamarneh, G. (2020). Deep semantic segmentation of natural and medical images: A review. *Artif. Intell. Review.* 54, 137–178. doi: 10.1007/s10462-020-09854-1

Torralba, A., Russell, B. C., and Yuen, J. (2010). LabelMe: Online image annotation and applications. *Proc. IEEE.* 98, 1467–1484. doi: 10.1109/JPROC.2010.2050290

Wang, H., Chen, X., Zhang, T., Xu, Z., and Li, J. (2022). CCTNet: Coupled CNN and transformer network for crop segmentation of remote sensing images. *Remote. Sens.* 14, 1956. doi: 10.3390/rs14091956

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in Proceedings of the European conference on computer vision (ECCV) (Berlin: Springer). 3–19. doi: 10.1007/978-3-030-01234-2_1

Wu, S., Sun, Y., and Huang, H. (2021). "Multi-granularity feature extraction based on vision transformer for tomato leaf disease recognition," in 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST) (Guangzhou). 387–390. doi: 10.1109/iaecst54258.2021.9695688

Yang, K., Zhong, W., and Li, F. (2020). Leaf segmentation and classification with a complicated background using deep learning. *Agronomy.* 10, 1721. doi: 10.3390/agronomy10111721

Yan, Q., Yang, B., Wang, W., Wang, B., Chen, P., and Zhang, J. (2020). Apple leaf diseases recognition based on an improved convolutional neural network. *Sensors (Basel Switzerland).* 20, 3535. doi: 10.3390/s20123535

Ye, W. (2014). Cotton breeding research progress in China. *New Biotechnol.* 31, 168. doi: 10.1016/J.NBT.2014.05.2038

Yuan, H., Zhu, J., Wang, Q., Cheng, M., and Cai, Z. (2022). An improved DeepLab v3+ deep learning network applied to the segmentation of grape leaf black rot spots. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.795410

Yu, C., Wang, J., Gao, C., Yu, G., Shen, C., and Sang, N. (2020). "Context prior for scene segmentation," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (Piscataway, NJ: Computer Vision Foundation / IEEE). 12413–12422. doi: 10.1109/cvpr42600.2020.01243

Zhang, X., Yang, Q., Zhou, R., Zheng, J., Feng, Y., Zhang, B., et al. (2022). Perennial cotton ratoon cultivation: A sustainable method for cotton production and breeding. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.882610

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Los Alamitos,CA: IEEE Computer Society). 6230–6239. doi: 10.1109/CVPR.2017.660

Zoph, B., and Le, Q. V. (2017). "Neural architecture search with reinforcement learning," in International Conference on Learning Representations (ICLR) 2017 (Ithaca, NY: OpenReview.net). doi: 10.48550/arXiv.1611.01578

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 31 January 2023
DOI 10.3389/fpls.2023.1101143

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Rajasheker Reddy Pullanagari,
Massey University, New Zealand
Brandon Weihs,
United States Department of Agriculture
(USDA), United States

*CORRESPONDENCE
Jiangwei Qiao
qiaojiangwei@caas.cn

SPECIALTY SECTION
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

RECEIVED 17 November 2022
ACCEPTED 11 January 2023
PUBLISHED 31 January 2023

CITATION
Li J, Li Y, Qiao J, Li L, Wang X, Yao J and
Liao G (2023) Automatic counting of
rapeseed inflorescences using deep
learning method and UAV RGB imagery.
Front. Plant Sci. 14:1101143.
doi: 10.3389/fpls.2023.1101143

COPYRIGHT
© 2023 Li, Li, Qiao, Li, Wang, Yao and Liao.
This is an open-access article distributed
under the terms of the Creative Commons
Attribution License (CC BY). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Automatic counting of rapeseed inflorescences using deep learning method and UAV RGB imagery

Jie Li[1], Yi Li[1], Jiangwei Qiao[2]*, Li Li[3], Xinfa Wang[2], Jian Yao[3] and Guisheng Liao[4]

[1]Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan, China, [2]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China, [3]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, [4]National Lab of Radar Signal Processing, Xidian University, Xi'an, China

Flowering is a crucial developing stage for rapeseed (*Brassica napus* L.) plants. Flowers develop on the main and branch inflorescences of rapeseed plants and then grow into siliques. The seed yield of rapeseed heavily depends on the total flower numbers per area throughout the whole flowering period. The number of rapeseed inflorescences can reflect the richness of rapeseed flowers and provide useful information for yield prediction. To count rapeseed inflorescences automatically, we transferred the counting problem to a detection task. Then, we developed a low-cost approach for counting rapeseed inflorescences using YOLOv5 with the Convolutional Block Attention Module (CBAM) based on unmanned aerial vehicle (UAV) Red–Green–Blue (RGB) imagery. Moreover, we constructed a Rapeseed Inflorescence Benchmark (RIB) to verify the effectiveness of our model. The RIB dataset captured by DJI Phantom 4 Pro V2.0, including 165 plot images and 60,000 manual labels, is to be released. Experimental results showed that indicators R^2 for counting and the mean Average Precision (mAP) for location were over 0.96 and 92%, respectively. Compared with Faster R-CNN, YOLOv4, CenterNet, and TasselNetV2+, the proposed method achieved state-of-the-art counting performance on RIB and had advantages in location accuracy. The counting results revealed a quantitative dynamic change in the number of rapeseed inflorescences in the time dimension. Furthermore, a significant positive correlation between the actual crop yield and the automatically obtained rapeseed inflorescence total number on a field plot level was identified. Thus, a set of UAV-assisted methods for better determination of the flower richness was developed, which can greatly support the breeding of high-yield rapeseed varieties.

KEYWORDS

rapeseed, deep learning, unmanned aerial vehicle, attention mechanism, rapeseed inflorescence counting, seed yield

1 Introduction

Rapeseed (*Brassica napus* L.) is one of the main oil crops, and the development of the rapeseed industry is very important to secure oil supply. The Chinese rapeseed oilseed crop accounts for around 20% of world production. According to the yearbook of the China National Bureau of Statistics, the rapeseed planting area has decreased from 7,316,000 ha in 2010 to 6,765,000 ha in 2020, but the yield per unit area and the total output rose to 2,077 kg/ha and 14.094 million tons in 2020, compared with 1,748 kg/ha and 12.788 million tons in 2010, respectively (National Bureau of Statistics of China, 2021), which shows that the Chinese rapeseed industry has made great achievements in recent years. Yet, rapeseed output is comparatively low considering the consumption for more than 10 consecutive years, as the demand for rapeseed oil increases with the development of the economy and the continuous improvement of consumption level. Affected by the dual effects of rural labor transfer and the shock of imported oil crops, domestic industry development is under enormous pressure (Wang, 2018; Liu et al., 2019). Although China is a major producer of rapeseed oil and rapeseed meal, it mainly relies on inventories and imports to make up for the supply gap (He et al., 2022). Breeding rape varieties with higher oil yield is essential.

Field-based phenotyping plays a vital role in the process of plant breeding for plant performance evaluation (Yang et al., 2017; Jin et al., 2021). The flowering stage of rapeseed lasts as long as 30 days, accounting for almost one-fourth of the growth period. It is a critical period for breeders to analyze the factors that affect the seed yield. The yield components of rape include the number of seeds per pod, the pods, and the weight of each seed (Tayo and Morgan, 1975; Mcgregor, 1981). The number of pods retained at maturity has the greatest effect on the seed yield (Tayo and Morgan, 1975; Mcgregor, 1981; Diepenbrock, 2000; Sonja et al., 2007; Faraji, 2010; Gan et al., 2016; Kirkegaard et al., 2018), which is largely decided by flowering time and flower production that can have potential to turn into pods (Tayo and Morgan, 1975; Faraji et al., 2008; Faraji, 2010; Gong et al., 2018; Kirkegaard et al., 2018; Zhang and Flottmann, 2018; Matar et al., 2020; Zhang et al., 2021). Meanwhile, the amount of rapeseed flowers is closely related to the number of rapeseed inflorescences. As a result, counting rapeseed inflorescences is essential. It is of great significance to explore the correlation between the total number of rapeseed inflorescences of each plot and the seed yield in order to improve the yield of rapeseed. Previous studies highlight the relevance of flowering dates (Han et al., 2021), peak flowering (D'Andrimont et al., 2020), and coverage (Zhang et al., 2021) in seed yield and quality. There is a lack of quantitative description of the number of rapeseed inflorescences. A huge number of rapeseed inflorescences make manual counting impossible. Therefore, an automatic, rapid, and non-destructive method of rapeseed inflorescence counting is useful in plant breeding.

Nowadays, application of satellite remote sensing technology in agriculture has become a trend (Marie et al., 2020). This method obtains multispectral crop information in different periods and spaces from a large area without destroying the crop structure of crops. As a consequence, it has been widely used in precision agriculture (Jin et al., 2017; Khanal et al., 2017; Yang, 2020), yield prediction (Arab et al., 2021; He et al., 2021; Shuai and Basso, 2022), etc. Satellite images have also been successfully applied to rapeseed monitoring. D'Andrimont et al. (2020) combined optical and radar images captured by the Copernicus Sentinel-1 and Sentinel-2 satellite sensors to estimate the flowering timing. Han et al. (2021) utilized the Landsat-8 and Sentinel-1/2 selected from Google Earth to monitor flowering traits. However, low spatial resolution and the period limited the development of satellite images in precision agriculture. Flowering time may be missed because the phenotype of flowers in image changes significantly over a long period of time. Additionally, the size of rapeseed inflorescence is relatively tiny, making it difficult to count tiny objects from a low- resolution satellite image. Recently, the rapid development of unmanned aerial vehicle (UAV) technology provides a new opportunity for continuous acquisition of rapeseed data under different growing stages and it supports different image resolutions. The flexibility and convenience establish an easy way to monitor flowering crops (Kumar et al., 2021; Xu et al., 2021). Wan et al. (2018) employed Red–Green–Blue (RGB) and multispectral images to establish a model to estimate yellow flower number. Zhang et al. (2021) indicated that normalized difference yellowness index-based flowering pixel numbers could estimate flowering intensity by UAV. Sun et al. (2022) extracted spectral traits and structural traits to simultaneously predict wheat yield and grain protein content by multispectral and LiDAR data. Indeed, a lot of plant information is obtained from multispectral data, which is helpful for phenotype analysis. However, multispectral acquisition is greatly affected by weather. The flowering time of different rapeseed materials is inconsistent, and the state of rapeseed flowers changes rapidly in a field. When capturing spectral data, the weather conditions need to be kept as constant as possible and the reliance on it limits multispectral acquisition schemes. It is better to have a UAV scheme that is less affected by the weather.

UAV equipped with RGB cameras (UAV-RGB) has the advantages of higher resolution and less weather affection, which makes it possible to acquire and process large-scale field information conveniently. Recently, combined with the deep learning technology, UAV-RGB system extracted purple leaves (Zhang et al., 2020a), recognized frozen (Li et al., 2022), and estimated stand count (Zhang et al., 2020b) of rapeseed effectively. Many experts are also attracted to study crop counting. Plant counting is regarded as an object- counting task in the computer vision area (Lu and Cao, 2020). For example, Liu et al. (2020a) counted rice to estimate density using the deep learning method. However, these methods discard location information of the plant and the poor explainability limits the counting performances (Lu et al., 2022). More researchers deal with counting as an object detection task, in which target quantity could be estimated from the number of detected bounding boxes. These methods are proven to outperform some traditional machine learning models in counting of maize (Kumar et al., 2021), cotton bloom (Xu et al., 2018), sorghum heads and wheat ears (Lu and Cao, 2020; Lu et al., 2022), etc. Nevertheless, only little attention pays on automatic counting of the rapeseed inflorescences using UAV-RGB because counting rapeseed inflorescences is a challenging task. Inflorescence varies in climate, cultivars, and agricultural management, whose shapes have different degrees of adhesion and occlusion in RGB images. Furthermore, the rape common data set is deficient due to the limitation of rapeseed growth area and time-consuming data annotation, which results in insufficient model training and generates significant challenges to accurate counting.

In order to count rapeseed inflorescences accurately and quickly, we treat the counting task as a detection task. Object detection is a

fundamental task in the computer vision area. Current state-of-the-art object detectors are generally divided into two categories, namely, one-stage and two-stage. Faster R-CNN (Ren et al., 2017) is a traditional two-stage network, which proves to be suitable for various plants and plant-organ detection (Madec et al., 2019; Liu et al., 2020b; Xu et al., 2021). These networks have the property of proposal optimization mechanism. Consequently, two-stage operation provides a high accuracy but slow speed and poor real-time performance, which is difficult to meet the requirements of high-throughput and efficient detection. Taking detection as a regression process, the one-stage object detection network does not select candidate regions separately but omits the candidate region generation step. Instead, it integrates feature extraction, target classification, and position regression into one stage for operation. This single- stage network is represented by SSD (Liu et al., 2016), RetinaNet (Lin et al., 2020), and YOLO series (Redmon et al., 2016; Bochkovskiy et al., 2020; Ultralytics, 2021). It has the potential to be faster and simpler (Marie et al., 2020; Yang et al., 2021) but has trailed the accuracy of two-stage detectors (Subramanian and Selvi, 2021). YOLOv5 (Ultralytics, 2021), as an one-stage classic deep recognition end-to-end network model, is the latest version of YOLO series. The model improves the detection speed while maintaining the detection accuracy of existing models. It is one of the optimal choices for the high-throughput detection. Consequently, the improved YOLOv5 model is proposed to enhance the effect of the model in detecting rapeseed inflorescences in the state of dense adhesive occlusion.

This article aims to automatically quantify the total number of rapeseed inflorescences of each plot precisely and quickly. To our knowledge, this is the first time that the deep learning method and UAV-RGB system have been combined to count rapeseed inflorescences. Furthermore, we investigate the correlation between the number of rapeseed inflorescence and the seed yield based on the proposed method. The contributions of this paper are summarized as follows:

- We transformed the rapeseed inflorescences counting as a detection problem and developed an improved YOLOv5 model.
- We built a novel Rapeseed Inflorescence Benchmark (RIB), containing 165 plot images with 60,000 manual labels.
- We assessed the accuracy and robustness of the proposed deep learning algorithms from different rapeseed inflorescences densities, sites, and years.
- We analyzed the correlation between the number of the rapeseed inflorescences and yield and employed the application of inflorescence number in breeding.

2 Platform and data preparation

2.1 Study area

Rapeseed is divided into winter rapeseed (planted at the end of September and harvested in May of the following year) and spring rapeseed (planted at the end of April and harvested in September). Between them, the planting area and output of winter rapeseed account for more than 90% of the country and one-fourth of the world, mainly located in the Yangtze River Basin region. Our study areas belonged to these area. It is located at Yangluo Base (114.51409E, 30.71047N, at an altitude of 24 m, subtropical monsoon climate) of Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, Hubei Province, China. The details are shown in Figure 1. Two experimental fields (rapeseed field A and rapeseed field B) were presented in this study area, of which 252 and 165 plots were chosen, respectively. The training data were composed of the RGB images obtained from rapeseed field B, from which our counting model was obtained. The testing data, consisting of the images acquired from rapeseed field A, were used for testing the robustness of the model with 24 materials planted. In field A and field B, there were two types of plots of different sizes in each field, 8.0 m^2 (2.0 m long × 4.0 m wide) and 6.0 m^2 (2 m long × 3 m wide). Experiments were carried out during the flowering stage, from February to May in 2021 and 2022.

2.2 Data acquisition

The original image data were collected by DJI Phantom 4 Pro V2.0 [1] equipped with an RGB camera. This UAV-RGB system was a consumer-grade drone with a 20- MP (5,472 × 3,648 pixels) image resolution. Rapeseed images were acquired once a week under sunny and partly cloudy conditions in the noon hours (11:00–13:00), local time. Details of the data acquisition environment are shown in Table 1. The UAV was set in an automatic acquisition mode to capture images at the speed of 1.9 m/s and a course overlap rate of 75%. Image data acquisition in the study area was completed within 2 h. To obtain images of different scales, we took flight altitudes of 10 and 15 m.

2.3 Image dataset

RGB images captured from the UAV were stitched and calibrated automatically by Agisoft PhotoScan [2]. Thereafter, we obtained six orthophotos of experimental field A and four orthophotos of experimental field B in different periods in 2021. In order to verify the robustness of the model, we acquired a digital orthophoto map of the field in 2022, which would also be applied for further analysis. The deep learning method was a data-driven technique; thus, a large number of samples were needed to train representatives of the detection network model. Rapeseed inflorescence appeared in various forms because of the different flowering time, weather condition, flower size, location, cultivars, postures, colors, occlusions and adhesions, etc. Therefore, enough samples from different flowering stages and conditions were needed to prepare to extract the robust feature.

Details of the construction of the Rapeseed Inflorescence Benchmark (RIB) were as follows. Adobe Photoshop was exploited to crop the field orthophoto into plot sizes. The cropped image resolution was 1,900 × 1,600 for the 8.0- m plot and 1,820 × 680 pixels for the 6.0-m^2 plot. The RIB dataset included 165 different plot images from field B. The total number of rapeseed inflorescences of each plot varied from

1 https://www.dji.com/cn/phantom-4-pro-v2

2 https://www.agisoft.com/

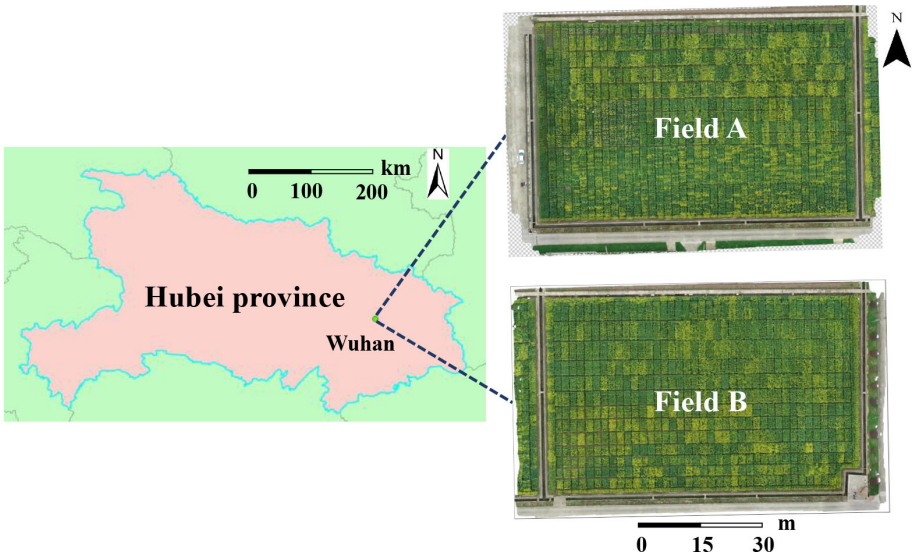


FIGURE 1
The study area (114.51409E, 30.71047N) at Yangluo Base of Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China.

TABLE 1 Image acquisition at Wuhan, Hubei Province, China, in 2021 and 2022.

Year	Acquisition dates	Flight altitude	Field	Temperature	Environment
	February 19	15m	Field A	12°C	
	February 26	15m	Fields A and B	14°C	
2021	March 3	10m	Fields A and B	21°C	11:00 to 13:00, low/middle cloudy or cloudless, wind speed less than 4 m/s, good light conditions
	March 14	10m	Fields A and B	23°C	
	March 22	10m	Fields A and B	24°C	
	April 5	10m	Field A	15°C	
2022	March 1	13m	Field B	20°C	

0 to 1,200. Each rapeseed inflorescence was manually labeled with a rectangle using the LabelImg toolbox [3]. Finally, 60,000 labels were obtained as ground truth for our experiment.

3 Materials and methods

In order to address the problem of rapeseed inflorescence counting automatically and efficiently, we proposed a count approach based on deep learning using UAV-RGB imagery. This paper defined a counting object from a top-down perspective. The whole workflow of the proposed approach is presented in Figure 2. Firstly, field images were captured during the flowering stage by UAV-RGB imagery. Next, original images were spliced to generate an orthophoto map by Agisoft PhotoScan. Then, the field orthophoto map was cropped to produce plot images by Photoshop [4]. Afterward, the sample set consisting of all plot images was split to construct train and test datasets to train the network using YOLOv5 combined with the CBAM. After getting the counting results, this paper presented the number difference of rapeseed inflorescences in time series and in field. At last, we investigated the correlation between the total number of rapeseed inflorescences of each plot and the seed yield.

3.1 Deep learning method

3.1.1 YOLOv5 network

As a one-stage end-to-end detection network, YOLOv5 transferred detection as a regression process. The network structure consisted of three parts: backbone, neck, and head. First of all, the backbone part was a CSPDarknet53 network for feature extraction and included the structures of Focus, Convolution, Batch Normalization, and SiLU (CBS), C3, and Spatial Pyramid Pooling (SPP). The Focus structure was mainly used for slicing operation. In this way, the width and height information of the pictures was concentrated in the channel space, which could obtain a double

3 https://github.com/tzutalin/labelImg

4 https://www.adobe.com/cn/products/photoshop.html

10.3389/fpls.2023.1101143

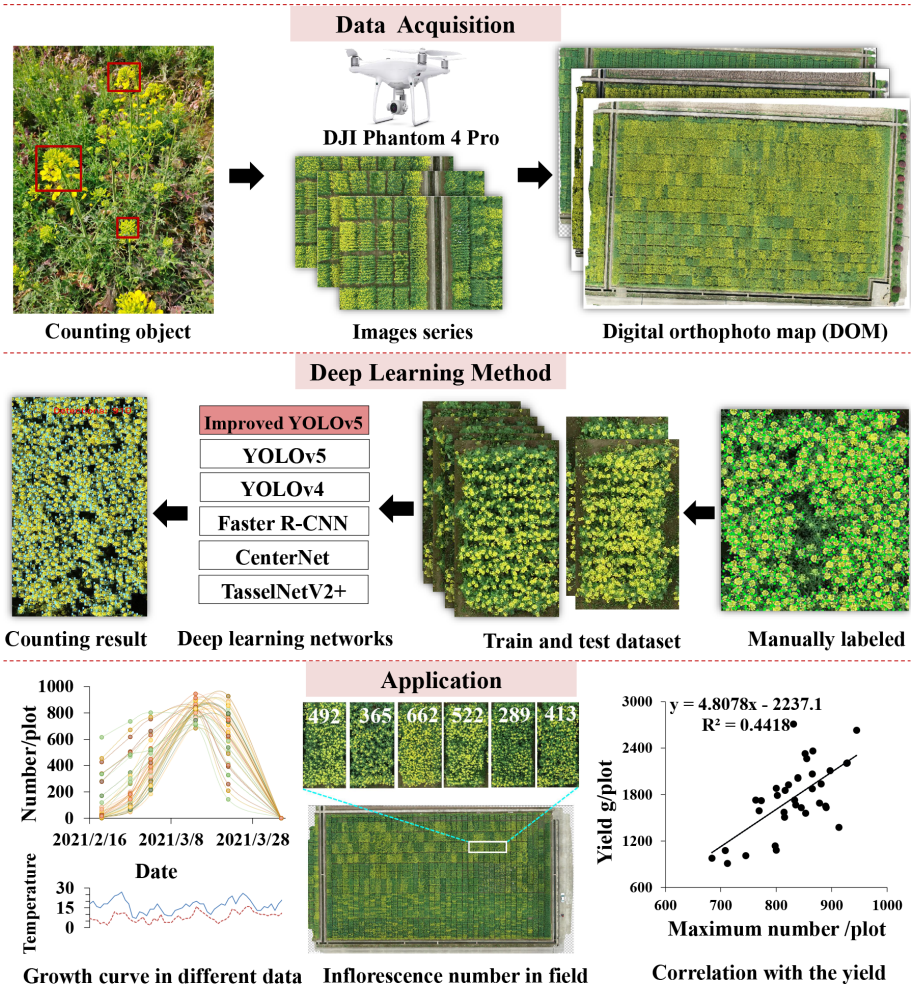


FIGURE 2
General frame diagram of this study.

downsampling feature map without information loss. The C3 structure is the combination of three convolutional blocks along with the CSP bottleneck. This module is the main module for learning the residual characteristics. The SPP part used different pool kernels to maximize the pooling of the feature map and subsequently spliced feature maps of different scales. This operation combined local features with global features and enriched the expression ability of the feature map. Accordingly, the feature extraction structure that was input with a picture would generate three kinds of scale feature maps. Next, the neck part utilized for feature fusion included Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). FPN fused features through top-down upsampling, whereas PAN transmitted features in a bottom-up pyramid, which enhanced the feature fusion ability of different layers and carried out multiscale prediction. Ultimately, consisting of bounding box loss (regression loss function) and non-maximum suppression (nms), the head structure predicted the category and the location of the object. In addition, YOLOv5 chose Generalized Intersection over Union Loss (GIoU Loss) as the loss function. Three scale feature maps were eventually generated to predict large, medium, and small targets, respectively, in the detection layer of YOLOv5.

YOLOv5 contained four different network structures, namely, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The depth and width of the network were controlled by different numbers of residual components in the CSP network, leading to four different structures of YOLOv5. With the increase in depth and width of the network, the ability of learning, feature extraction, and fusion also improved, whereas it paid the cost of longer computing time. Among them, YOLOv5x achieved the highest accuracy but the slowest detection speed. YOLOv5s had the fastest detection speed, three times that of YOLOv5x. As a result, YOLOv5s could better meet the real-time requirements whereas YOLOv5x was able to conform to the condition of high-precision detection.

3.1.2 The improved YOLOv5 network

Input with an image, YOLOv5 would generate bounding boxes for predicting object location in the network. Basically, the rapeseed inflorescence appeared tiny in UAV-RGB images and was in dense overlapping occlusion state. YOLOv5 had the ability to detect small objects. However, it was very difficult to extract enough features due to the variety of the rapeseed inflorescences, especially for adhesion and mutual occlusion. The attention mechanism was imitated from

the trait of human vision and widely used in the field of computer vision (Zhu et al., 2019; Hu et al., 2020; Wang et al., 2022a), because it enhanced the key information of the feature and improved the detection accuracy to a certain extent.

The CBAM is a lightweight attention module proposed by Woo et al. (2018), which can perform attention operations in spatial and channel dimensions. It consists of two independent sub-modules, named Channel Attention Module (CAM) and Spatial Attention Module (SAM). CAM learns the weights of different channels and then multiplies them with the weights to enhance the attention to the key channel domain. SAM focuses on the location of the image target. Miao et al. (2022) added the CBAM attention module to YOLOv5 to enhance the feature extraction ability of the backbone network to detect infrared ships. Hoang et al. (2022) utilized the YOLOv5 model with the CBAM attention module to identify various small targets in thermal images. Wang et al. (2022b) integrated the CBAM into the YOLO network and proposed a lightweight one-stage network called the Mobile Ghost Attention YOLO network to improve the performance of the model and the detection of apple leaf diseases. Ye et al. (2022) integrated Convolutional Block Attention Module (CBAM) and Efficient Channel Attention (ECA) into the neck of the latest YOLOv5 network to identify the terminal bud of Chinese fir seedlings in complex backgrounds. The applications mentioned above show that the CBAM attention module is extremely easy to integrate into the YOLOv5 network structure for better representation of objects by focusing on important features and neglecting unnecessary ones without adding too much complexity to the network (Hoang et al., 2022). Thus, to extract the characteristics of the dense rapeseed inflorescences fully, we embedded a CBAM attention mechanism to the YOLOv5 backbone network to attract its attention to local small target features in UAV images to detect dense rapeseed inflorescences well.

Different from the above improvement methods, we added the CBAM between C3 and CBS in the backbone. Figure 3 presents the detailed modification area of YOLOv5-CBAM. The Channel Attention Module of the CBAM enhanced the feature expression of the occluded target, and the Spatial Attention Module highlighted the detecting areas in the feature map. This operation improved the effectiveness and comprehensiveness of feature extraction. It was also convenient for more sufficiency in feature fusion. Moreover, this method worked for the four versions of YOLOv5, making the deployment of YOLOv5-CBAM flexible. The detection results of rapeseed inflorescences were presented in a detection bounding box, whereas the counting results were shown in the counting number.

3.2 Evaluation

For the interpretability of counting, this paper selected the detection method for counting. The counting results came from the number of detected bounding boxes. This method could determine the location of rapeseed inflorescence while counting

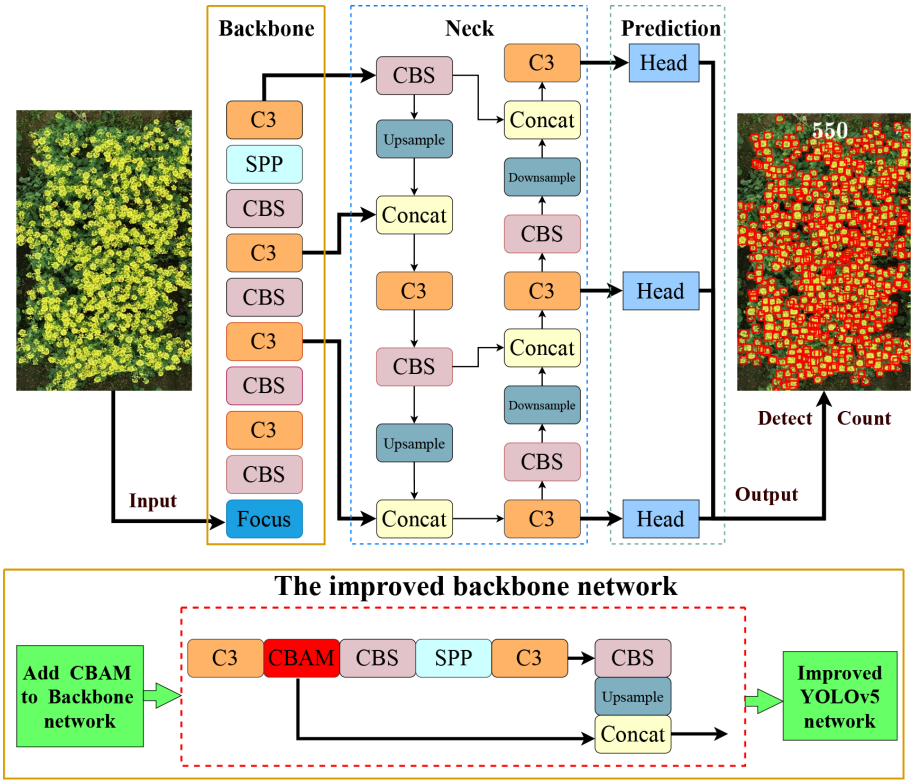


FIGURE 3
Network structure diagram of the proposed method. The upper part is the YOLOv5 model, and the lower part is the improved backbone network and composition of each module. The improved YOLOv5 generated detection map with the number of rapeseed inflorescences after inputting a rapeseed image.

and is helpful in analyzing the reasons for missing or error count. Thus, the accuracy of the counting was closely related to the detection performance.

3.2.1 Detection performance

The paper used Precision, Recall, F1-score, and mean Average Precision (*mAP*) to evaluate the detection performance. Precision is the ratio between the number of accurately predicted samples and the number of actually detected samples in the predicted samples; Recall is the ratio between the number of accurately predicted samples and the total number of samples.

F1-score is a comprehensive evaluation index that assesses the detection accuracy of the model and denotes the harmonic average of Precision and Recall. *mAP* is a measure of the performance of object detectors. The four evaluation metrics are defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$m\text{AP} = \frac{1}{C}\sum_{k=i}^{N} P(k)\Delta R(k) \tag{4}$$

where True Positive (TP) denotes the total number of rapeseed inflorescences correctly predicted in each plot, demonstrating that the correct detection result and the ground truth coincide. False Positive (FP) indicates the total number of rapeseed inflorescences that is incorrectly predicted in each plot, which means that the detection result is rapeseed inflorescence but the ground truth is not. False Negative (FN) is the number of non-flower samples that is incorrectly predicted as rapeseed inflorescences. C is the number of categories. Since we only need to detect one category, this value is set to 1. N represents the number of all pictures in the test set. $P(k)$ means the Precision when k pictures can be recognized, and $\Delta R(k)$ denotes the change of the Recall value when the number of recognized pictures changes from k-1 to k.

3.2.2 Counting performance

To verify the accuracy of the counting, the coefficient of determination (R^2) and the root mean square error (RMSE) are used as evaluation indicators to measure the counting performance of the model. The coefficient of determination means the degree of fit of the regression model and represents the correlation between the predicted value and the true value. The higher the value of R^2, the better the fitting effect of the model is. RMSE denotes the degree of deviation between the true number and the predicted number. The lower the value of RMSE, the better the counting performance of the model is. They are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(m_i - c_i)^2}{\sum_{i=1}^{n}(m_i - \overline{m})^2} \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - c_i)^2} \tag{6}$$

where n represents the number of plots, m_i and c_i represent the total number of rapeseed inflorescences manually labeled and counted by the model in the ith image, respectively, and $\overline{m}$ represents the average number of inflorescences manually labeled in all plots.

4 Results and discussion

4.1 Network performance

4.1.1 Training details

In this study, the operating system used was the Ubuntu 20.04 operating system with an NVIDIA GTX 3090 GPU. The implementation based on PyTorch 1.8.0 with torchvision 0.9.0. We divided the sample sets into training set, test set, and validation set according to 7:2:1. The initial learning rate and batch size were set to 0.01 and 8, respectively. In order to better optimize the objective function, cosine annealing was utilized to reduce the learning rate in the process of model training. With the increase in the number of iterations, the learning rate first decreased slowly, then accelerated, and then slowly decreased again with the change trend of cosine. Adam optimizer was exploited to train the detection methods. In the training process, the CIoU loss value of the model gradually decreased as the number of iterations increased. The validation loss values changed greatly at the beginning and fell gradually before 150 epochs. After around 400 epochs, both of the loss values tended to be stable and close to a convergent state. Trained for 400 epochs, the final network model of detection was obtained and utilized to detect and count rapeseed inflorescences.

4.1.2 Ablation study

To verify the effect of the CBAM attention mechanism on network detection, we performed ablation experiments in this part. Four network structures of YOLOv5 and the network model with the CBAM attention mechanism were tested on the RIB dataset. From the evaluation results in Table 2, we observed that the Precision of the four YOLOv5-CBAM models improved from 0.6% to 1.0% compared with the model without the CBAM. Notably, only the Recall of YOLOv5x-CBAM increased whereas the Recall of the other three YOLOv5 models with the CBAM attention mechanism decreased due to the interaction between Precision and Recall. Compared with the non-CBAM model, the F1-score values of YOLOv5m, YOLOv5l, and YOLOv5x with the CBAM were increased except for the decline of YOLOv5s-CBAM because these two indicators restricted each other in actual situations. Based on the above, *mAP* was exploited to evaluate the detection performance of the model.

10.3389/fpls.2023.1101143

TABLE 2 Experimental results of four YOLOv5 network structures before and after improvement.

Model	CBAM	Precision (%)	Recall (%)	F1-score (%)	mAP (%)	R^2	RMSE	Parameters	FPS (f/s)
YOLOv5s	–	90.4	84.8	87.5	90.0	0.943	68.1	**706,3542**	**83.8**
	√	91.4	82.9	86.9	90.9	0.960	56.8	707,4329	83.4
YOLOv5m	–	89.9	85.9	87.9	90.7	0.951	69.2	214,25046	64.8
	√	90.5	85.8	88.1	91.9	0.948	65.1	214,57729	61.7
YOLOv5l	–	91.2	84.7	87.8	90.1	0.962	55.4	472,86710	52.3
	√	**92.2**	83.9	87.9	91.6	0.964	54.3	473,38473	51.2
YOLOv5x	–	89.5	87.0	88.1	91.1	0.954	60.9	882,68374	44.1
	√	90.1	**87.4**	**88.7**	**93.6**	**0.966**	**52.1**	883,43313	44.8

Bold means best value.

After embedding the CBAM, the *mAP* of YOLOv5x, as the largest improvement in the four YOLOv5 structures, was up to 2.5%. For the counting indicators, R^2 varied from 0.943 to 0.966. Error and missing detection led to inaccuracy in this value because R^2 was an overall statistic. The RMSE decreased more than the model without the CBAM, showing that adding an attention mechanism to the traditional network was effective in improving the counting accuracy of rapeseed inflorescence. The model parameters increased by 10,787, 32,683, 51,763, and 74,940, respectively, but the FPS of the model with the CBAM changed slightly. These means that the target detection time remained basically unchanged after adding a lightweight attention module. From the above analysis, we concluded that it was feasible to count the rapeseed inflorescences using UAV images and deep learning.

For visual comparison, we utilized a similar image to visualize the detecting results in Figure 4. The left image and the subimage in Figure 4 were produced by YOLOv5x, and the right one was generated by YOLOv5x-CBAM. Because the flight height of the UAV was more than 10 m, rapeseed inflorescence presented different degrees of blur in the image. Under these circumstances, even though targets occluded and stuck, both methods could detect correctly most of rapeseed inflorescences. However, YOLOv5x had obvious missed detection in some cases shown in enlarged subimages with red arrows. After adding the attention mechanism, the model could distinguish the difference in adhesion of rapeseed inflorescences and reduce the possibly of missing detection.

4.2 Comparison with different networks

To verify the comprehensive performance of the improved network in detection and counting, a study was conducted to compare the performances of several other classical network models, including Faster R-CNN, YOLOv4, CenterNet, and TasselNetV2+. Faster R-CNN (Ren et al., 2017), a traditional two-stage detecting method, utilized the Region Proposal Network (RPN) to generate candidate regions and then combined candidate region generation, feature extraction, target classification, and position regression to achieve an end-to-end structure target detection model. YOLOv4 (Bochkovskiy et al., 2020), a one-stage detecting method and a state-of-the-art object detector, improved and optimized various parts of YOLO series before. Instead of detecting object bounding boxes, TasselNetV2+ (Lu and Cao, 2020) was

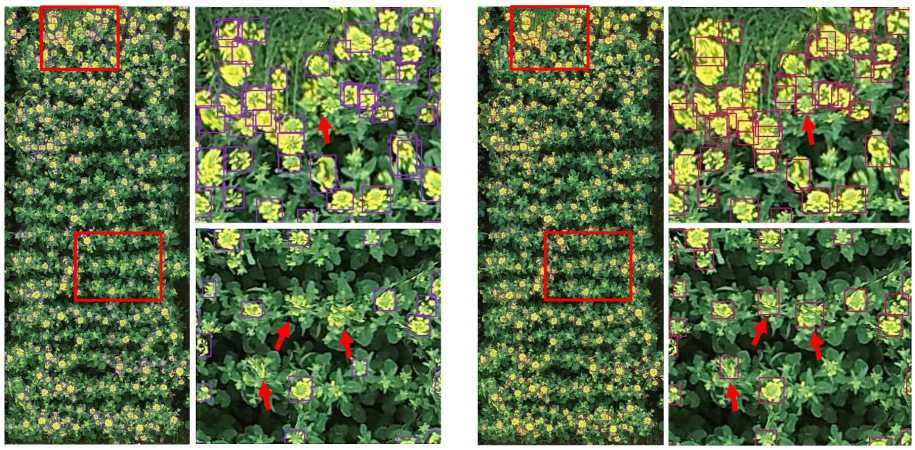


FIGURE 4
Comparison of detection results between YOLOv5x and YOLOv5x-CBAM. On the left is the YOLOv5x detection result and the partial enlarged view. On the right is the YOLOv5x-CBAM detection result and the partial enlarged view.

another plant counting method that directly regressed the number of objects in an image.

As shown in Table 3, quantitative results were obtained by using a default parameter setting. The results presented that the proposed approach achieved the best performance in F1-score and *mAP*, both of which were up to 88.7 and 93.6, respectively. Although CenterNet obtained the highest Precision and FPS, its F1-score and Recall were extremely low. It means that the detection speed of CenterNet was the fastest among the five networks, but it failed to correctly predict the rapeseed inflorescence as target in the case of more dense objects and led to a large number of false inspections. The R^2 and RMSE value of our method were 0.966 and 52.1, respectively. The counting performance was better than other classical networks. To further prove the validity of the counting, we conducted the experiment to explore the correlation between manual counts and network inferred counts. As shown in Figure 5, our method had a strong correlation between manual counts and network model inferred counts. Results in the sixth row of Table 3 presented that the proposed approach outperformed the state-of-the-art detection methods in RMSE. We could clearly see that our method provided

TABLE 3 Comparison of experimental results between the proposed method and other classical networks.

Methods	Precision (%)	Recall (%)	F1-score (%)	mAP (%)	R^2	RMSE	FPS (f/s)
Faster R-CNN	68.9	53.5	60.0	62.3	0.821	120.2	21.6
YOLOv4	94.0	74.2	83.0	91.4	0.926	180.9	40.5
CenterNet	**95.1**	61.3	75.0	86.3	0.964	53.9	**77.9**
TasselNetV2+	–	–	–	–	0.951	63.3	22.0
The proposed method	90.1	**87.4**	**88.7**	**93.6**	**0.966**	**52.1**	44.8

Bold means best value.

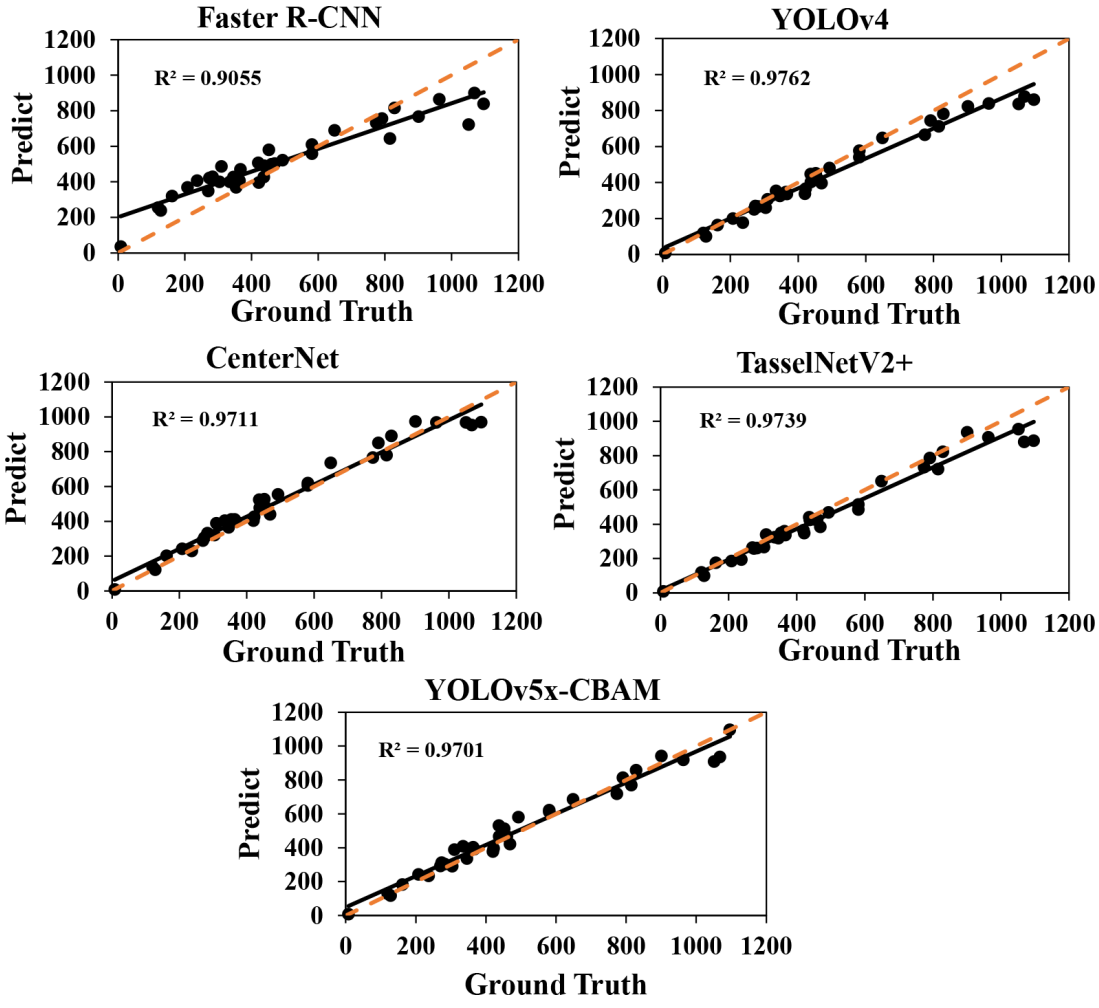


FIGURE 5
Fitting curves of predicted and true values of five network models. The R^2 value represents the fitting effect. The orange dotted line represents the 1: 1 fitted line.

not only interpretable counting results but also detection boxes, providing a basis for subsequent counting improvements.

To compare with the ground truth, shown in Figure 6, six heatmaps were produced to visualize the counting results. First of all, we output the coordinate value of the prediction box of each traditional network for the same image, then calculated the center point coordinate, and generated the heatmap through Gaussian blur. The heatmap of each traditional network denoted its counting results

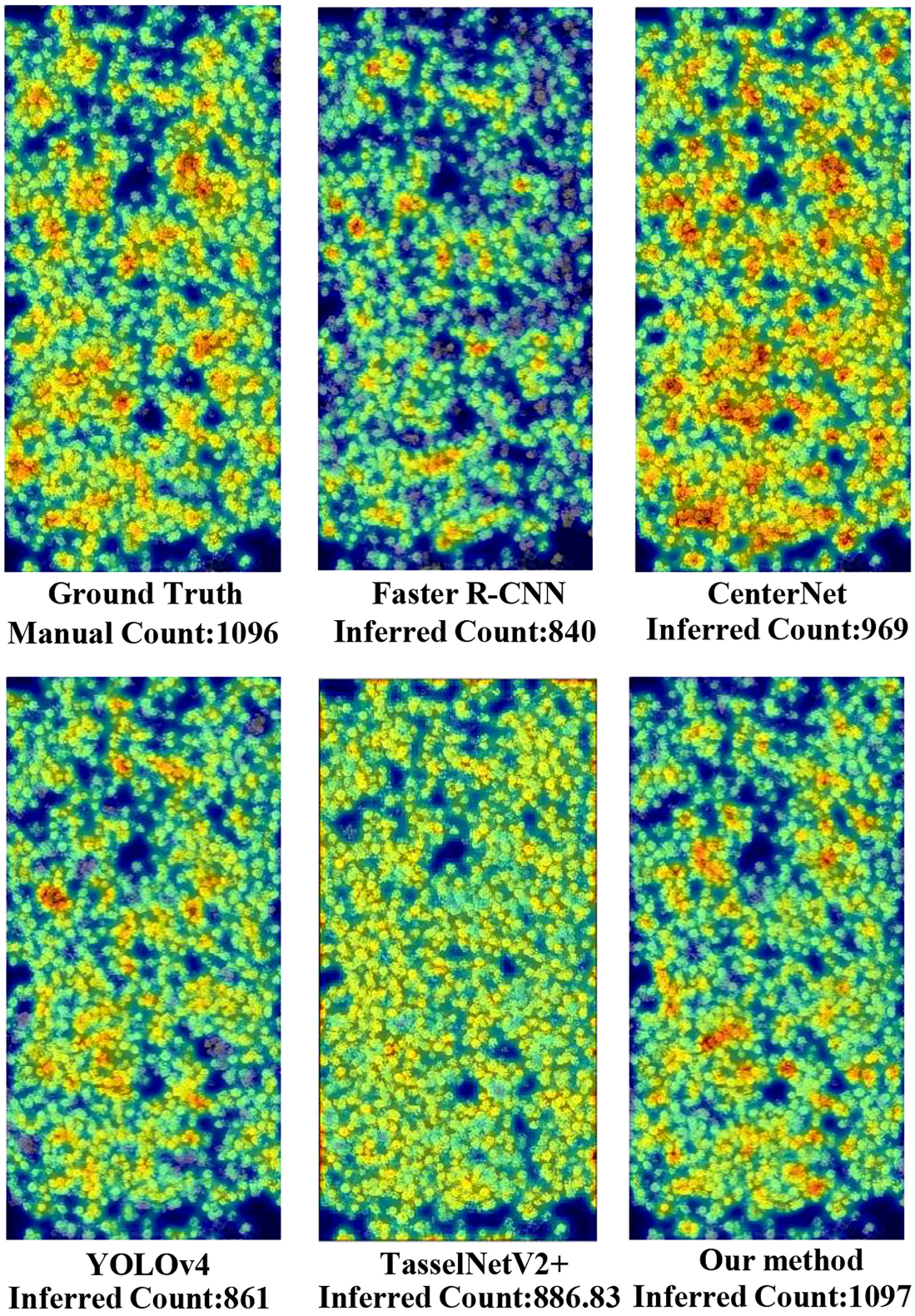


FIGURE 6
Comparison of visualization results between five networks and the ground truth.

and the density of rapeseed inflorescences. From the heatmaps, we observed that the detection results of YOLOv4 and the proposed method were closer to ground truth. However, YOLOv4 missed many adhesive rapeseed inflorescence targets, leading to low accuracy of the rapeseed inflorescences counting. For CenterNet, the total counting result was near the ground truth. However, it had duplicate count in the area with dense rapeseed inflorescences and missing count in other areas. This visual result was consistent with the low value of the F1-score indicator shown in Table 3. Compared with Faster R-CNN and TasselNetV2+, the detection and counting results of rapeseed inflorescences obtained with our method in the heat map were almost the same as those of the ground truth, indicating that our method was more accurate in detecting and counting rapeseed inflorescences.

4.3 The accuracy in different rapeseed densities

In order to further explore the factors affecting the accuracy of counting, we performed tests using three datasets with different rapeseed densities. As shown in Figure 7, we took images of sparse, medium-density, and dense rapeseed from the three datasets that contained 11, 13, and 12 plot images, respectively, as samples. It is observed that green leaves accounted for a large proportion in the sparse rapeseed inflorescence dataset. On the contrary, only yellow rapeseed inflorescences were seen in the dense rapeseed image dataset. The medium-density rapeseed inflorescence dataset contained at least one-third of the visible green leaves from the image and a part of blooming rapeseed.

Table 4 lists the evaluation results. The *mAP* in the three datasets were 79.6%, 91.5%, and 92.1% and the R^2 reached 0.888, 0.930, and 0.927, respectively, meaning that the method proposed by this paper could be applied at any flower stages. However, in the dense rapeseed inflorescence test dataset, the *mAP* decreased to 79.6%. This result further suggested the difficulty of detecting and counting rapeseed inflorescences in occlusion and adhesion areas.

4.4 Experiment on the new test site in 2022

To validate the robustness and effectiveness of our method, we conducted experiments at the two novel test sets that were called test

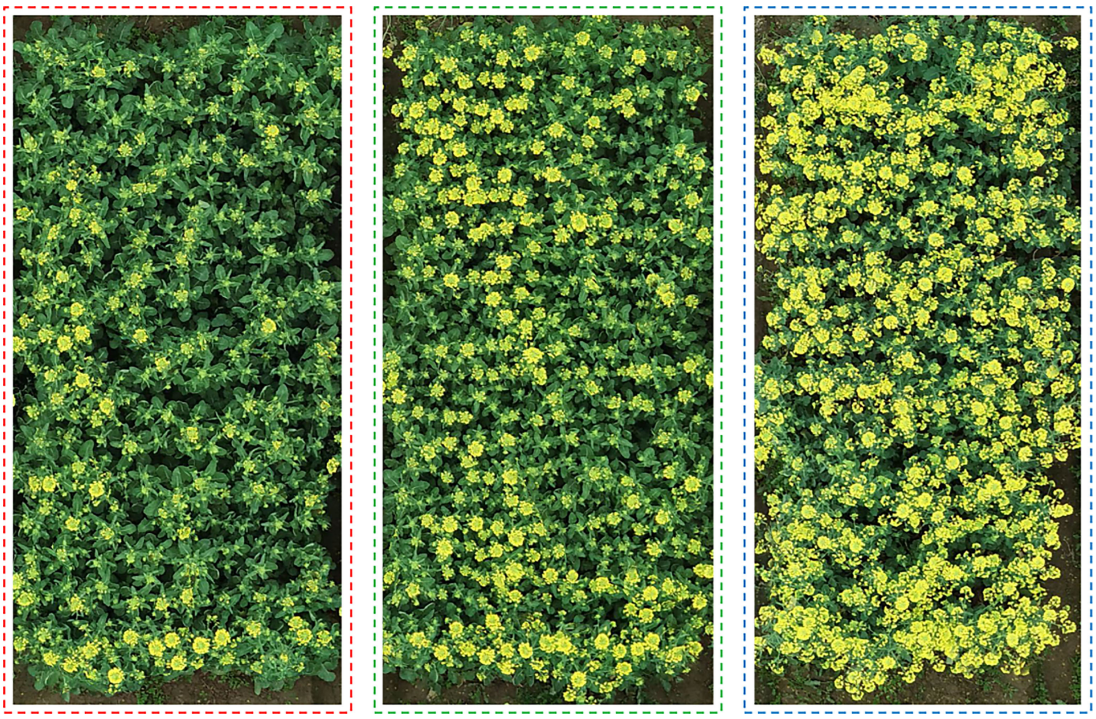


FIGURE 7
Distribution of rapeseed inflorescence under different densities. The figures on the left, middle, and right represent sparse rape inflorescence, medium-density, and dense rape inflorescences, respectively.

TABLE 4 Experimental results of the proposed method at different densities.

Density	Precision (%)	Recall (%)	F1-score (%)	mAP (%)	R^2	RMSE (%)	Speed (ms)
Dense	85.7	74.0	79.4	79.6	0.888	83.2	12.3
Middle	**92.2**	84.6	88.2	91.5	**0.930**	45.9	12.5
Sparse	89.4	**87.8**	**88.6**	**92.1**	0.927	**30.2**	**11.3**

Bold means best value.

10.3389/fpls.2023.1101143

set A and test set B with manually annotated bounding boxes. Test set A was randomly selected from field A in 2021. Test set B included 30 plot images at the early flowering stage, which we presented in Figure 8 with the red box area in the digital orthophoto maps. A strong correlation was found between manual counts and predicted counts on these two test sets. As shown in Figure 8, the value of R^2 on test sets A and B rose up to 0.96 and 0.97, respectively. Our method could well detect and count the rapeseed inflorescences not only in the year of experiment but in the years to come. Therefore, it is applicable to most scenarios of rapeseed inflorescence detection and counting and will meet the needs of conventional farmland management.

4.5 The application of counting number

4.5.1 The trend of the number of rapeseed inflorescences in the flowering period

The trend of the number change of rapeseed inflorescences is very important for breeding, because we can observe the flowering and withering times of materials. We chose 40 plots from field A randomly and counted the number of rapeseed inflorescences of each plot in six different periods by utilizing the proposed method. The data of six periods are shown in the x-axis, beginning in February 13, 2021, and ending in April 9, 2021. As shown in Figure 9A, the y-axis is the rapeseed inflorescence number. The range varied from 0 to 1000. Figure 9B presents the minimum and maximum temperatures of the flowering period in the study area. The maximum temperature during the day was around 20C, and the minimum temperature at night was around 5C, which was suitable for the growth of rapeseed. The paper exploited dotted lines of different colors to fit the change in the number of rapeseed inflorescences in different plots in the flowering period. From Figure 9A, we observe that most of the plots bloomed around February 18 and withered after March 20, and the peak flowering time was around mid and late March.

4.5.2 The differences between the quantity of the rapeseed inflorescences in field

In addition to observing one material in time order, we compared the differences between the quantity of the rapeseed inflorescences in field. One picture of test field A in March 3, 2021, was chosen. The number of rapeseed inflorescences in different plots was quickly and quantitatively obtained by the proposed method. The counting results are shown in Figure 10. The red box and its enlarged subimage included 30 plot images, and each number marked on these images represents its corresponding counting result. Compared with visual observation, the counting results were consistent with the growth trend of the rapeseed. We quantified the change of flowering stage by

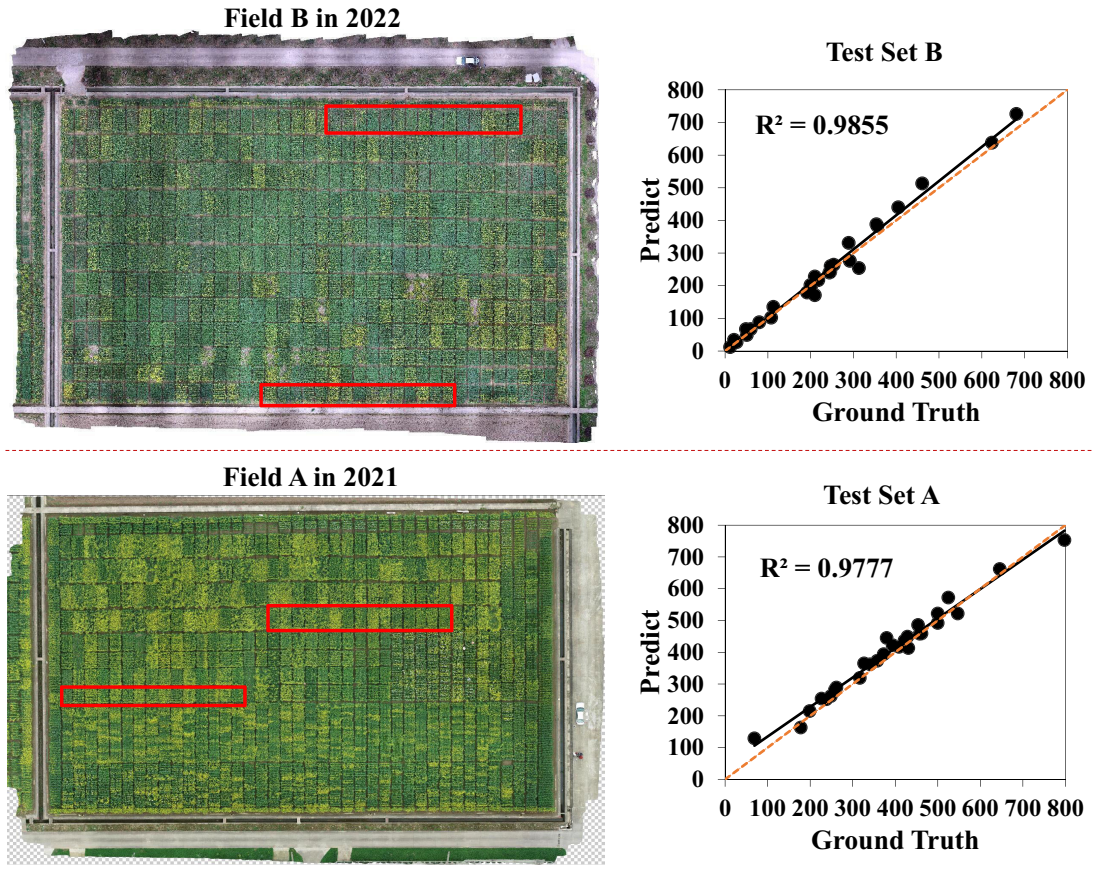


FIGURE 8
Comparison between inferred counts and manual counts on two novel test sets. Test set B from field B in 2022 is represented by the red box area in the figure's upper left corner, whereas test set A from field A in 2021 is represented by the red box area in the figure's lower left corner. The fitting curves of test set B and test set A between the predicted value and the ground truth obtained by the proposed approach in different fields and years are respectively shown on the right. The orange dotted line represents the 1: 1 fitted line.

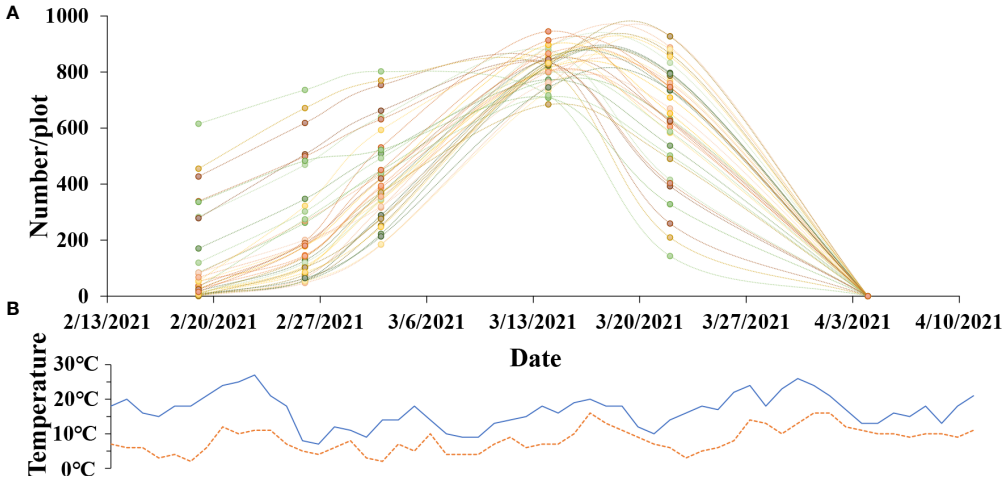


FIGURE 9
The growth of rapeseed progress during the flowering stage. **(A)** Change curve of the total number of rapeseed inflorescences of each plot. **(B)** Fluctuation curve of maximum temperature and minimum temperature.

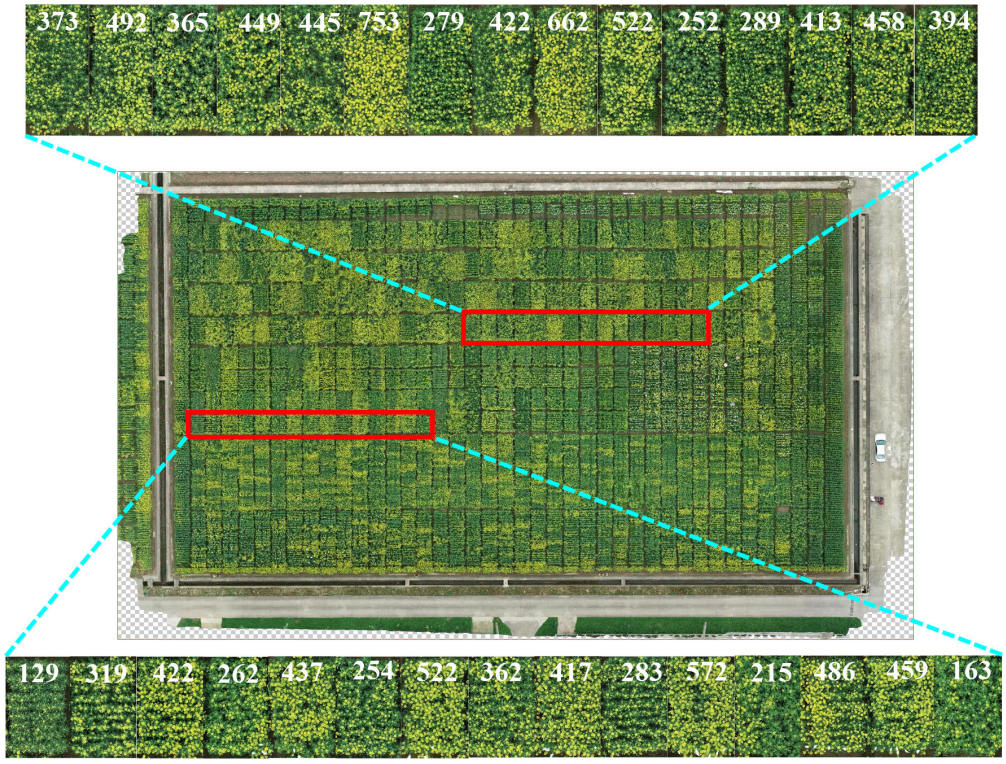


FIGURE 10
Experimental results of 30 test plots randomly selected in the new field obtained by the proposed method during the flowering stage in field A, March 3, 2021. The red box represents test plots, and the enlarged subimage represents count results.

automatic counting, which brought convenience for breeders to analyze the material performance in the flowering period.

4.5.3 The number of rapeseed inflorescences and the seed yield

The number of rapeseed inflorescences in the flowering period is an important predictor for its seed yield. To further study the effect of rapeseed inflorescence number on yield, we analyzed the correlation between the number of rapeseed inflorescences and the seed yield. We used the data from five flowering periods of field A. The total number of rapeseed inflorescences predicted by the proposed network model for each plot and the corresponding seed yield were recorded to explore the correlation between them. The coefficient of determination (R^2) was exploited to reflect the fitting degree of the linear regression model, representing the interpretation degree of the total number of rapeseed inflorescences of each plot to the seed yield.

10.3389/fpls.2023.1101143

Initially, we utilized some representative plots that included two different area sizes to perform a regression analysis between the sum and max of the total number of rapeseed inflorescences of each plot in the five periods and the seed yield. As shown in Figure 11, the R^2 value in Figures 11A, B reached 0.2714 and 0.3874, respectively. The R^2 of the sum of the rapeseed inflorescences in the five periods was slightly lower than that of the max of the rapeseed inflorescences in the five periods. It might be that we collected data at a low frequency. There was a sequence in flowers blooming, different materials, and planting methods leading to different flowering times that led to the omission of many rapeseed inflorescences in the whole flowering period.

Although flower formation at different stages contributed less or more, they might still have the potential to affect the seed yield. Consequently, some plots with equal plant area, good growth, and relatively high yield were selected to further investigate the correlation between the maximum of the total number of rapeseed inflorescences of each plot in the five periods and the seed yield. The results are shown in Figure 12. The R^2 value, reaching 0.4418, indicated that there was a significant correlation between the maximum of the total number of the rapeseed inflorescences of each plot in the five periods and the seed yield.

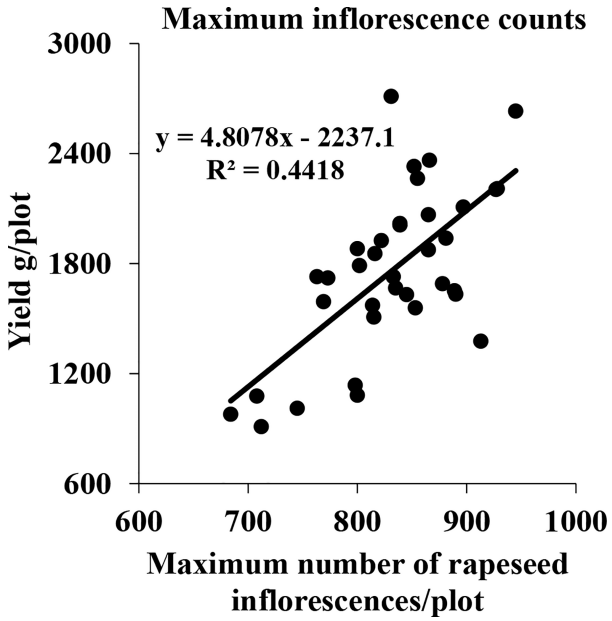


FIGURE 12
Correlation between the seed yield and maximum number of all rapeseed inflorescences of each plot in five periods in field A in 2021.

5 Conclusion

The flowering stage of rapeseed is a critical period for breeders to analyze the factors that affect the seed yield. In this paper, we apply the YOLOv5-CBAM method to quantify the total number of rapeseed inflorescences of each plot automatically, precisely, and quickly. The results show that the detecting precision is up to 91.7% in the RIB dataset. Additionally, we verify the robustness of the proposed method in the datasets of sparse, medium-density, and dense rapeseed inflorescences and found that our method is suitable for rapeseed inflorescences with different densities. Moreover, we conduct comparative experiments on several classical counting networks, including YOLOv4 for one-stage detection networks, Faster R-CNN for two-stage networks, CenterNet for a detection method without an anchor box, and TasselnetV2+ for a counting method without detection boxes. The experimental results verify the effectiveness of the proposed method. In fact, due to its strong robustness and effectiveness, our method suits most scenarios and can also be applied to the detection and counting of other crops that are manually labeled, such as apple, wheat, cotton, and sunflower.

However, inflorescences on the same rapeseed plant have different performances in different periods. In this paper, only several periods of rapeseed are selected as representatives, and a problem of insufficient sample performance appears. Subsequent studies will consider sample diversities and enrich the benchmark. On the other

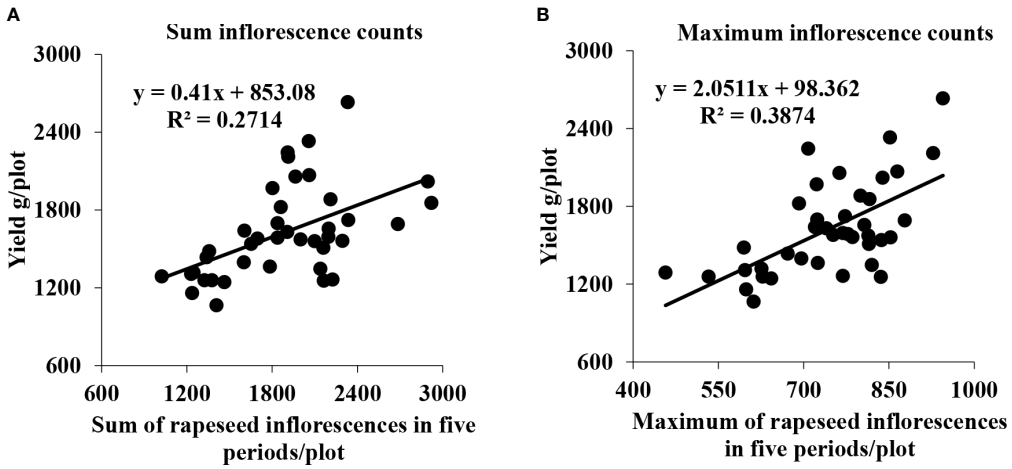


FIGURE 11
Scatter plots between the seed yield and the total number of rapeseed inflorescences of the representative plots. **(A)** Correlation between the seed yield and sum of all rapeseed inflorescences of each plot in five periods. **(B)** Correlation between the seed yield and maximum count of all rapeseed inflorescences of each plot in five periods.

hand, the accuracy of detection directly affects the effect of counting. In the case of dense rapeseed inflorescences, the detecting accuracy is only 85.7%, showing that a lot of room for improvement still remains to be done in this case, in which we will further consider designing more attention modules for the rapeseed inflorescence detection to increase the accuracy.

Finally, we perform regression analysis between the total number of rapeseed inflorescences of each plot and the seed yield to find out their correlation. Under the multiperiod factor, there is a significant correlation between the maximum of the total number of rapeseed inflorescences of each plot in five periods and the seed yield. The total number of rapeseed inflorescences obtained with our method in each plot is the primary factor that affects the yield. However, the seed yield is also affected by many other factors, such as seed weight, seed number per silique, photosynthetic capacity of silique wall, and the abiotic resistance of plants. Thus, we will focus on the pod stage of rapeseed in the future. Lodging after flowering and UAV combined with multi-spectrum are considered to estimate pods. Then, combining the different factors was done to conduct multiple regression analysis to predict yield more accurately.

Based on the nature of the rapeseed growth process, the number of rapeseed inflorescences gradually increases until the flowers withered. During this period, the number of rapeseed inflorescences is closely related to the seed yield. However, we can only capture images of the rapeseed inflorescences in six different periods of flowering. Considering the insufficient data of the whole flowering period, we will increase the sampling frequency in flowering period to better fit the change curve of the number of rapeseed inflorescences in future work.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/LYLWYH/Rapeseed-Data.

Author contributions

All authors made significant contributions to this manuscript. JL, YL, and JQ performed field data collection and wrote the manuscript. JQ and LL designed the experiment. JY, XW, and GL provided suggestions on the experiment design. All authors read and approved the final manuscript.

Funding

This work was supported by the Agricultural Science and Technology Innovation Project (CAAS-ZDRW202105), the National Natural Science Foundation of China (Project No. 32172101), and the Open Foundation of Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System(Project No. HBSEES202206).

Acknowledgments

A large part of the code is borrowed from https://github.com/ultralytics/yolov5. Thanks are given for their wonderful works.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Arab, S. T., Noguchi, R., Matsushita, S., and Ahamed, T. (2021). Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach. *Remote Sens. Applications: Soc. Environ.* 22, 100485. doi: 10.1016/j.rsase.2021.100485

Bochkovskiy, A., Wang, C., and Liao, H. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv* 2004, 10934. doi: 10.48550/arXiv.2004.10934

D'Andrimont, R., Taymans, M., Lemoine, G., Ceglar, A., Yordanov, M., and Velde, M. V. D. (2020). Detecting flowering phenology in oil seed rape parcels with sentinel-1 and -2 time series. *Remote Sens. Environ.* 239, 111660. doi: 10.1016/j.rse.2020.111660

Diepenbrock, W. (2000). Yield analysis of winter oilseed rape (Brassica napus l.): a review. *Field Crops Res.* 67, 35–49. doi: 10.1016/S0378-4290(00)00082-4

Faraji, A. (2010). Flower formation and pod/flower ratio in canola (Brassica napus l.) affected by assimilates supply around flowering. *Int. J. Plant Production* 4, 271–280. doi: 10.22069/IJPP.2012.710

Faraji, A., Latifi, N., Soltani, A., and Rad, A. H. S. (2008). Effect of high temperature stress and supplemental irrigation on flower and pod formation in two canola (Brassica napus l.) cultivars at mediterranean climate. *Asian J. Plant Sci.* 7, 343–351. doi: 10.3923/ajps.2008.343.351

Gan, Y., Harker, K. N., Kutcher, H. R., Gulden, R. H., Irvine, B., May, W. E., et al. (2016). Canola seed yield and phenological responses to plant density. *Can. J. Plant Sci.* 96, 151–159. doi: 10.1139/cjps-2015-0093

Gong, Y., Duan, B., Fang, S., Zhu, R., Wu, X., Ma, Y., et al. (2018). Remote estimation of rapeseed yield with unmanned aerial vehicle (UAV) imaging and spectral mixture analysis. *Plant Methods* 14, 70. doi: 10.1186/s13007-018-0338-z

Han, J., Zhang, Z., and Cao, J. (2021). Developing a new method to identify flowering dynamics of rapeseed using landsat 8 and sentinel-1/2. *Remote Sens.* 13, 105. doi: 10.3390/rs13010105

He, W., Li, J., Wang, X., Lin, Q., and Yang, X. (2022). Current status of global rapeseed industry and problems, countermeasures of rapeseed industry in china. *China Oils Fats* 47, 1–7. doi: 10.19902/j.cnki.zgyz.1003-7969.210430

He, X., Yang, L., Li, A., Zhang, L., Shen, F., Cai, Y., et al. (2021). Soil organic carbon prediction using phenological parameters and remote sensing variables generated from sentinel-2 images. *Catena* 205, 105442. doi: 10.1016/j.catena.2021.105442

Hoang, V. P., Ninh, H., and Hai, T. T. (2022). "CBAM-YOLOv5 for infrared image object detection," in *Artificial intelligence and machine learning in defense applications IV* (Berlin, Germany: SPIE), 116–127. doi: 10.1117/12.2640690

Hu, J., Shen, L., Samuel, A., Sun, G., and Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023. doi: 10.1109/TPAMI.2019.2913372

Jin, X., Liu, S., Baret, F., Hemerlé, M., and Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sens. Environ.* 198, 105–114. doi: 10.1016/j.rse.2017.06.007

Jin, X., Zarco-Tejada, P. J., Schmidhalter, U., Reynolds, M. P., Hawkesford, M. J., Varshney, R. K., et al. (2021). High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. *IEEE Geosci. Remote Sens.* 9, 200–231. doi: 10.1109/MGRS.2020.2998816

Khanal, S., Fulton, J., and Shearer, S. (2017). An overview of current and potential applications of thermal remote sensing in precision agriculture. *Comput. Electron. Agric.* 139, 22–32. doi: 10.1016/j.compag.2017.05.001

Kirkegaard, J. A., Lilley, J. M., Brill, R. D., Ware, A. H., and Walela, C. K. (2018). The critical period for yield and quality determination in canola (Brassica napus l.). *Field Crops Res.* 222, 180–188. doi: 10.1016/j.fcr.2018.03.018

Kumar, A., Desai, S. V., Balasubramanian, V. N., Rajalakshmi, P., Guo, W., Balaji Naik, B., et al. (2021). Efficient maize tassel-detection method using UAV based remote sensing. *Remote Sens. Applications: Soc. Environ.* 23, 100549. doi: 10.1016/j.rsase.2021.100549

Li, L., Qiao, J., Yao, J., Li, J., and Li, L. (2022). Automatic freezing-tolerant rapeseed material recognition using uav images and deep learning. *Plant Methods* 18, 1–13. doi: 10.1186/s13007-022-00838-6

Lin, T., Goyal, P., Girshick, R., He, K., and Dollar, P. (2020). Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 318–327. doi: 10.1109/TPAMI.2018.2858826

Liu, C., Feng, Z., Xiao, T., Ma, X., Zhou, G., Huang, F., et al. (2019). Development, potential and adaptation of chinese rapeseed industry. *Chin. J. Oil Crop Sci.* 41, 485–489. doi: 10.7505/j.issn.1007-9084.2019.04.001

Liu, L., Lu, H., Li, Y., and Cao, Z. (2020a). High-throughput rice density estimation from transplantation to tillering stages using deep networks. *Plant Phenomics* 2020, 1375957. doi: 10.34133/2020/1375957

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016). "SSD: Single shot multibox detector," in *European Conference on computer vision* (Las Vegas, NV, USA: Springer), 21–37. doi: 10.1007/978-3-319-46448-0\s\do5(2

Liu, Y., Cen, C., Che, Y., Ke, R., and Ma, Y. (2020b). Detection of maize tassels from uav rgb imagery with faster r-CNN. *Remote Sens.* 12, 338. doi: 10.3390/rs12020338

Lu, H., and Cao, Z. (2020). Tasselnetv2+: A fast implementation for high-throughput plant counting from high-resolution RGB imagery. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.541960

Lu, H., Liu, L., Li, Y., Zhao, X., Wang, X., and Cao, Z. (2022). Tasselnetv3: Explainable plant counting with guided upsampling and background suppression. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3058962

Madec, S., Jin, X., Lu, H., Solan, B. D., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorology* 264, 225–234. doi: 10.1016/j.agrformet.2018.10.013

Marie, W., Frédéric, J., and Grgory, D. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402. doi: 10.1016/j.rse.2019.111402

Matar, S., Kumar, A., Holtgräwe, D., Weisshaar, B., and Melzer, S. (2020). The transition to flowering in winter rapeseed during vernalization. *Plant Cell Environ.* 44, 506–518. doi: 10.1111/pce.13946

Mcgregor, D. I. (1981). Pattern of flower and pod development in rapeseed. *Can. J. Plant Sci.* 61, 275–282. doi: 10.4141/cjps81-040

Miao, L., Li, N., Zhou, M., and Zhou, H. (2022). "CBAM-Yolov5: improved yolov5 based on attention model for infrared ship detection," in *International conference on computer graphics, artificial intelligence, and data processing (ICCAID 2021)* (Guangzhou, China: SPIE), 564–571. doi: 10.1117/12.2631130

National Bureau of Statistics of China (2021). *China Statistical yearbook-2021* (Bei Jing: China Statistics Press).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (Amsterdam, The Netherlands: IEEE), 779–788. doi: 10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Shuai, G., and Basso, B. (2022). Subfield maize yield prediction improves when in-season crop water deficit is included in remote sensing imagery-based models. *Remote Sens. Environ.* 272, 112938. doi: 10.1016/j.rse.2022.112938

Sonja, I., Cane, S., Zoran, D., Ana, M. J., Mirjana, J., and Ljupčo, J. (2007). Interrelationship between yield and yield related traits of spring canola (Brassica napus l.) genotypes. *Genetika* 39, 325–332. doi: 10.2298/GENSR0703325I

Subramanian, P., and Selvi, S. T. (2021). Detection of maturity stages of coconuts in complex background using faster r-CNN model. *Biosyst. Eng.* 202, 119–132. doi: 10.1016/j.biosystemseng.2020.12.002

Sun, Z., Li, Q., Jin, S., Song, Y., Xu, S., Wang, X., et al. (2022). Simultaneous prediction of wheat yield and grain protein content using multitask deep learning from time-series proximal sensing. *Plant Phenomics* 2022, 9757948. doi: 10.34133/2022/9757948

Tayo, T. O., and Morgan, D. G. (1975). Quantitative analysis of the growth, development and distribution of flowers and pods in oil seed rape (Brassica napus l.). *J. Agric. Sci.* 85, 103–110. doi: 10.1017/S0021859600053466

Ultralytics (2021). Available at: https://github.com/ultralytics/yolov5.

Wan, L., Li, Y., Cen, H., Zhu, J., Yin, W., Wu, W., et al. (2018). Combining uav-based vegetation indices and image classification to estimate flower number in oilseed rape. *Remote Sens.* 10, 10127–10134. doi: 10.3390/rs10091484

Wang, G., Zhai, Q., and Liu, H. (2022a). Cross self-attention network for 3d point cloud. *Knowledge-Based Syst.* 247, 108769. doi: 10.1016/j.knosys.2022.108769

Wang, H. (2018). New-demand oriented oilseed rape industry developing strategy. *Chin. J. Oil Crop Sci.* 40, 613. doi: 10.7505/j.issn.1007-9084.2018.05.001

Wang, Y., Wang, Y., and Zhao, J. (2022b). MGA-YOLO: A lightweight one-stage network for apple leaf disease detection. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.927424

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "CBAM: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)* (Munich, Germany: Springer International Publishing), 3–19. doi: 10.1007/978-3-030-01234-2\s\do5(1

Xu, R., Li, C., Paterson, A. H., Jiang, Y., Sun, S., and Robertson, J. S. (2018). Aerial images and convolutional neural network for cotton bloom detection. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02235

Xu, R., Li, C., Paterson, A. H., Jiang, Y., Sun, S., and Robertson, J. S. (2021). Estimates of maize plant density from uav rgb images using faster-RCNN detection model: impact of the spatial resolution. *Plant Phenomics* 3, 181–196. doi: 10.34133/2021/9824843

Yang, B., Gao, Z., Gao, Y., and Zhu, Y. (2021). Rapid detection and counting of wheat ears in the field using YOLOv4 with attention module. *Agronomy* 11, 10127–10134. doi: 10.3390/agronomy11061202

Yang, C. (2020). Remote sensing and precision agriculture technologies for crop disease detection and management with a practical application example. *Engineering* 6, 528–532. doi: 10.1016/j.eng.2019.10.015

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01111

Ye, Z., Guo, Q., Wei, J., Zhang, J., Zhang, H., Bian, L., et al. (2022). Recognition of terminal buds of densely-planted chinese fir seedlings using improved yolov5 by integrating attention mechanism. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.991929

Zhang, H., and Flottmann, S. (2018). Source-sink manipulations indicate seed yield in canola is limited by source availability. *Eur. J. Agron.* 96, 70–76. doi: 10.1016/j.eja.2018.03.005

Zhang, J., Xie, T., Yang, C., Song, H., Jiang, Z., Zhou, G., et al. (2020a). Segmenting purple rapeseed leaves in the field from UAV RGB imagery using deep learning as an auxiliary means for nitrogen stress detection. *Remote Sens.* 12, 1403. doi: 10.3390/rs12091403

Zhang, J., Zhao, B., Yang, C., Shi, Y., Liao, Q., Zhou, G., et al. (2020b). Rapeseed stand count estimation at leaf development stages with uav imagery and convolutional neural networks. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00617

Zhang, T., Vail, S., Duddu, H. S. N., Parkin, I. A. P., Guo, X., Johnson, E. N., et al. (2021). Phenotyping flowering in canola (Brassica napus l.) and estimating seed yield using an unmanned aerial vehicle-based imagery. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.686332

Zhu, X., Cheng, D., Zhang, Z., Lin, S., and Dai, J. (2019). "An empirical study of spatial attention mechanisms in deep networks," in *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)* (Seoul, South Korea: IEEE), 6688–6697. doi: 10.1109/ICCV.2019.00679

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 03 February 2023
DOI 10.3389/fpls.2023.1084778

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Marcin Wozniak,
Silesian University of Technology, Poland
Li Chaorong,
Yibin University, China

*CORRESPONDENCE
Sruti Das Choudhury
s.d.choudhury@unl.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 30 October 2022
ACCEPTED 11 January 2023
PUBLISHED 03 February 2023

CITATION
Das A, Das Choudhury S, Das AK, Samal A and Awada T (2023) EmergeNet: A novel deep-learning based ensemble segmentation model for emergence timing detection of coleoptile. *Front. Plant Sci.* 14:1084778. doi: 10.3389/fpls.2023.1084778

COPYRIGHT
© 2023 Das, Das Choudhury, Das, Samal and Awada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

EmergeNet: A novel deep-learning based ensemble segmentation model for emergence timing detection of coleoptile

Aankit Das[1†], Sruti Das Choudhury[2,3*†], Amit Kumar Das[4], Ashok Samal[1,2] and Tala Awada[3,5]

[1]Institute of Radio Physics and Electronics, University of Calcutta, Kolkata, West Bengal, India, [2]School of Computing, University of Nebraska-Lincoln, Lincoln, NE, United States, [3]School of Natural Resources University of Nebraska-Lincoln, Lincoln, NE, United States, [4]Department of Computer Science and Engineering, Institute of Engineering and Management, Kolkata, West Bengal, India, [5]Agricultural Research Division, University of Nebraska-Lincoln, Lincoln, NE, United States

The emergence timing of a plant, i.e., the time at which the plant is first visible from the surface of the soil, is an important phenotypic event and is an indicator of the successful establishment and growth of a plant. The paper introduces a novel deep-learning based model called EmergeNet with a customized loss function that adapts to plant growth for coleoptile (a rigid plant tissue that encloses the first leaves of a seedling) emergence timing detection. It can also track its growth from a time-lapse sequence of images with cluttered backgrounds and extreme variations in illumination. EmergeNet is a novel ensemble segmentation model that integrates three different but promising networks, namely, SEResNet, InceptionV3, and VGG19, in the encoder part of its base model, which is the UNet model. EmergeNet can correctly detect the coleoptile at its first emergence when it is tiny and therefore barely visible on the soil surface. The performance of EmergeNet is evaluated using a benchmark dataset called the University of Nebraska-Lincoln Maize Emergence Dataset (UNL-MED). It contains top-view time-lapse images of maize coleoptiles starting before the occurrence of their emergence and continuing until they are about one inch tall. EmergeNet detects the emergence timing with 100% accuracy compared with human-annotated ground-truth. Furthermore, it significantly outperforms UNet by generating very high-quality segmented masks of the coleoptiles in both natural light and dark environmental conditions.

KEYWORDS

event-based plant phenotyping, deep-learning, ensemble segmentation, emergence time detection, benchmark dataset

1 Introduction

Image-based plant phenotyping has the potential to transform the field of agriculture through the automated measurements of phenotypic expressions, i.e., observable biophysical traits of a plant as a result of complex interactions between genetics and environmental conditions. Accurate computation of meaningful phenotypes contributes to the study of high yield of better-quality crops with minimum resources (Das Choudhury et al., 2018). A plant's phenome is defined as its observable characteristics or traits and is determined by the complex interaction between genotype and the environment. Plant phenotyping analysis has been an active research field for some time that adds to the understanding of yield and resource acquisition, and therefore, accelerates breedingcycles, improves our understanding of plant responses to environmental stresses, and contributes to global food security under changing climate. Image-based plant phenotypes can be broadly classified into three categories: structural, physiological, and event-based (Das Choudhury et al., 2019). The structural phenotypes characterize a plant's morphology, whereas physiological phenotypes refer to the physiological processes that regulate plant growth and metabolism (Das Choudhury and Samal, 2020).

The timing detection of important events in a plant's life cycle, for example, the emergence of coleoptile (i.e., protective sheath covering the emerging shoot) and new leaves, flowering, and fruiting, from time-lapse sequences has recently drawn significant research attention. Such phenotypes are called event-based phenotypes and provide crucial information in understanding the plant's vigor, which varies with the interaction between genotype and environment. While interest in event-based phenotyping forleaves, flowers, and fruits has increased substantially in recent times (Wang et al., 2019; Bashyam et al., 2021), detecting the emergence and monitoring of the growth of the coleoptile based on computer vision and artificial intelligence techniques is a budding research field with vast opportunities for exploration. Emergence is a significant phenotype that not only helps determine the dormancy of seeds for different genotypes in different environmental conditions but also various aspects of early plant growth stages.

Unlike the visual tracking of rigid bodies, for instance, vehicles and pedestrians, the emergence timing detection of living organs and tracking their growth over time requires a different problem formulation with an entirely new set of challenges. Firstly, the state-of-the-art rigid body object detection and tracking methods deal with objects of considerably larger size that do not change in shape and appearance during the period of consideration. In contrast, our problem is to detect the coleoptile at emergence, when it is tiny in appearance, and track its dynamics as leaves emerge and grow into a seedling. The growth monitoring of size and shape is obtained as a by-product of an ensemble segmentation technique that segments the coleoptile with high accuracy. Secondly, the background (soil) in typical emergence detection in high-throughput plant phenotyping systems is significantly more complex than the state-of-the-art visual tracking applications. The soil substrate is multicolored due to the presence of perlite and vermiculite which makes the background cluttered, rendering the detection of a tiny coleoptile extremely challenging. Finally, the images are captured for a longer time than visual tracking, typically days, in a greenhouse with natural and artificial lighting conditions resulting in significant illumination variations.

The central contribution of this paper is to introduce a novel ensemble segmentation model tailored to the detection and growth monitoring of living organs in cluttered backgrounds and illumination variations for applications in event-based plant phenotyping. EmergeNet, characterized by its custom-designed loss function, uses a novel weighted ensemble learning technique to minimize the variance of the predicted masks and the generalization error for emergence timing detection. A benchmark dataset is indispensable for the development of the algorithm and performance comparison. Therefore, we have developed a publicly available benchmark dataset called the University of Nebraska-Lincoln Maize Emergence Dataset (UNL-MED) consisting of time-lapse image sequences of maize coleoptiles under the aforementioned conditions.

2 Related works

Multiple object tracking is challenging, yet it is of fundamental importance for many real-life practical applications (Xing et al., 2011). The survey paper by (Dhaka et al., 2021) provided comparisons of various convolutional neural networks and optimization techniques that are applied to predict plant diseases from leaf images. A comprehensive survey of multiple object tracking methods based on deep-learning is provided by (Xu et al., 2019). The method in (Xing et al., 2011) uses a progressive observation model followed by a dual-mode two-way Bayesian inference-based tracking strategy to track multiple highly interactive players with an abrupt view and pose variations in different sports videos, e.g., football, basketball, as well as hockey. A. Yilmaz et al (Yilmaz et al., 2006) showed that a plethora of research had been done in the field of object detection and tracking using various methods, including deep-learning algorithms. The method by (Aggarwal and Cai, 1999) gave an overview of the tasks involved in the motion analysis of a human body. (Dollár et al., 2009) worked on pedestrian detection, a key problem in computer vision, and proposed improved evaluation metrics. Computer vision based vehicle detection and tracking play an important role in the intelligent transport system (Min et al., 2018). The method in (Min et al., 2018) presents an improved ViBe for accurate detection of vehicles and uses two classifiers, i.e., support vector machine and convolutional neural network, to track vehicles in the presence of occlusions.

However, the use of deep neural networks for event-based plant phenotyping is in the early stage of research. The MangoYOLO algorithm (Wang et al., 2019) uses the YOLO object detector for detecting, tracking, and counting mangoes from a time-lapse video sequence. The method uses the Hungarian algorithm to correlate fruits between neighboring frames and a Kalman filter to predict the position of fruits in the following frames. A method for plant emergence detection and growth monitoring of the coleoptile based on adaptive hierarchical segmentation and optical flow using spatio-temporal image sequence analysis is presented in (Agarwal, 2017). A notable study in this domain includes the detection of budding and bifurcation events from 4D point clouds using a forward-backward analysis framework (Li et al., 2013). For

a large-scale phenotypic experiment, the seeds are usually sown in smaller pots until germination and then transplanted to bigger pots based on a visual inspection of the germination date, size, and health of the seedlings. The method by (Scharr et al., 2020) developed an image-based automated germination detection system based on transfer-learning deep neural networks equipped with a visual support system for inspecting and transplanting seedlings. Deep-learning based ensemble segmentation technique has been recently introduced in medical image processing in ratio-based sampling for the arteries and veins in abdominal CT scans (Golla et al., 2020), skin lesion diagnosis using dermoscopic images (Arulmurugan et al., 2021), and portrait segmentation for application in surveillance systems (Kim et al., 2021).

To the best of our knowledge, there is no previous research that accurately detects the emergence timing of seedlings from a cluttered soil background and tracks its growth over a time-lapse sequence under extreme variations in illuminations using deep-learning based ensemble segmentation with custom loss functions. This paper proposes a novel algorithm that not only detects the emergence of the coleoptile under all the aforementioned challenging conditions but also successfully tracks its growth by creating an overlay mask on the image sequence even under extremely low light conditions at night. The proposed model, EmergeNet, uses deep-learning algorithms to create a novel segmentation model which can predict segmentation masks with high accuracy. We also release a benchmark dataset with ground-truth called UNL-MED, consisting of 3832 high-definition time-lapse image sequences of the maize coleoptiles.

3 Materials and methods

3.1 Dataset description

Benchmark datasets are critical in developing new algorithms and performing uniform comparisons among state-of-the-art algorithms. Hence, we created a benchmark dataset called the UNL-MED. It is organized into two folders, namely, 'Dataset' and 'Training'. The 'Dataset' folder contains all of the 3832 raw high-definition images of resolution 5184 × 3456. The 'Training' folder contains two subfolders, namely, 'images', which has randomly selected 130 images for training, and 'masks', which contains 130 corresponding masks but downsampled to a resolution of 256 × 256. The images are captured at an interval of two minutes under various external conditions, including varying illumination, cluttered background, warm and cool tone, starting from before the emergence occurred until the coleoptile is about 1 inch high to facilitate its growth monitoring. Figure 1 is used to demonstrate an example of extreme contrast of illumination of the images used in the experiment based on histogram analysis. Figure 1A shows one of the brightest images and its histogram, whereas Figure 1B shows one of the darkest images and its corresponding histogram. Figure 1C shows the darkest image after histogram equalization and its histogram. Each image contains nine pots sowed with maize seeds of different genotypes. A visible light camera fitted with a tripod was placed directly above the nursery to capture high-definition top-view images of all nine pots every two minutes. The dataset can be freely downloaded from https://plantvision.unl.edu/dataset.

3.2 Dataset pre-processing

One of the most challenging and tedious tasks in image segmentation using deep-learning is the generation of ground-truth. In our case, it is in the form of binary masks corresponding to the plants in the images. For a custom dataset like the one used in this work, it is imperative that the masks are generated accurately, and therefore, it needs to be done manually. Utmost care has been taken while generating these masks as these serve as monitoring information during the semantic segmentation training to provide feedback to the neural network. In our experiment, we made use of the open-source manual annotation software developed by Visual Geometry Group (VGG) (Dutta et al., 2016). The flowchart of the data pre-processing is shown in Figure 2.

From this figure, we can see that each image is first hand-annotated, and then the corresponding data is exported in 'JSON' format for further processing. For each image, one or more masks are created from the exported data, and then they are superimposed to create the binary mask of the image. The images, along with their corresponding masks, are then fed into EmergeNet for training.

3.3 Proposed method: EmergeNet

In this section, we discuss the proposed model and its constituent parts in detail. Figure 3 represents the block diagram of the proposed method.

The first step is to generate masks from raw input images and pre-process them for training and evaluation. The images and their corresponding pixel labels are partitioned for training and testing. The training dataset is augmented to reduce overfitting and then fed to the model, EmergeNet, for training while the test dataset is used for evaluating the performance of the model. It can then be used to predict the emergence time of an image sequence. EmergeNet is a custom-made ensemble segmentation model that uses a weighted combination of loss functions and is specifically designed to detect tiny coleoptiles at the time of their first emergence from the soil under challenging conditions. It can also be used for growth monitoring of the plant over a time-lapse image sequence. EmergeNet consists of three underlying backbone architectures. The following subsections discuss these three standard backbone models, the EmergeNet Loss function, and the ensembling technique.

3.3.1 The backbone architectures

EmergeNet is built by ensembling three pre-trained networks, each based on the UNet architecture but with modified backbone models. It uses a custom-made loss function as well. The three backbone models used are SEResNet, InceptionV3, and VGG19.

3.3.1.1 The UNet architecture

The UNet architecture, which is an extension of an encoder-decoder convolutional network, is known for its precise segmentations using fewer training images. Therefore, it is only logical to make optimum utilization of the UNet architecture for the task of fine-grain semantic segmentation. The basic intuition

10.3389/fpls.2023.1084778

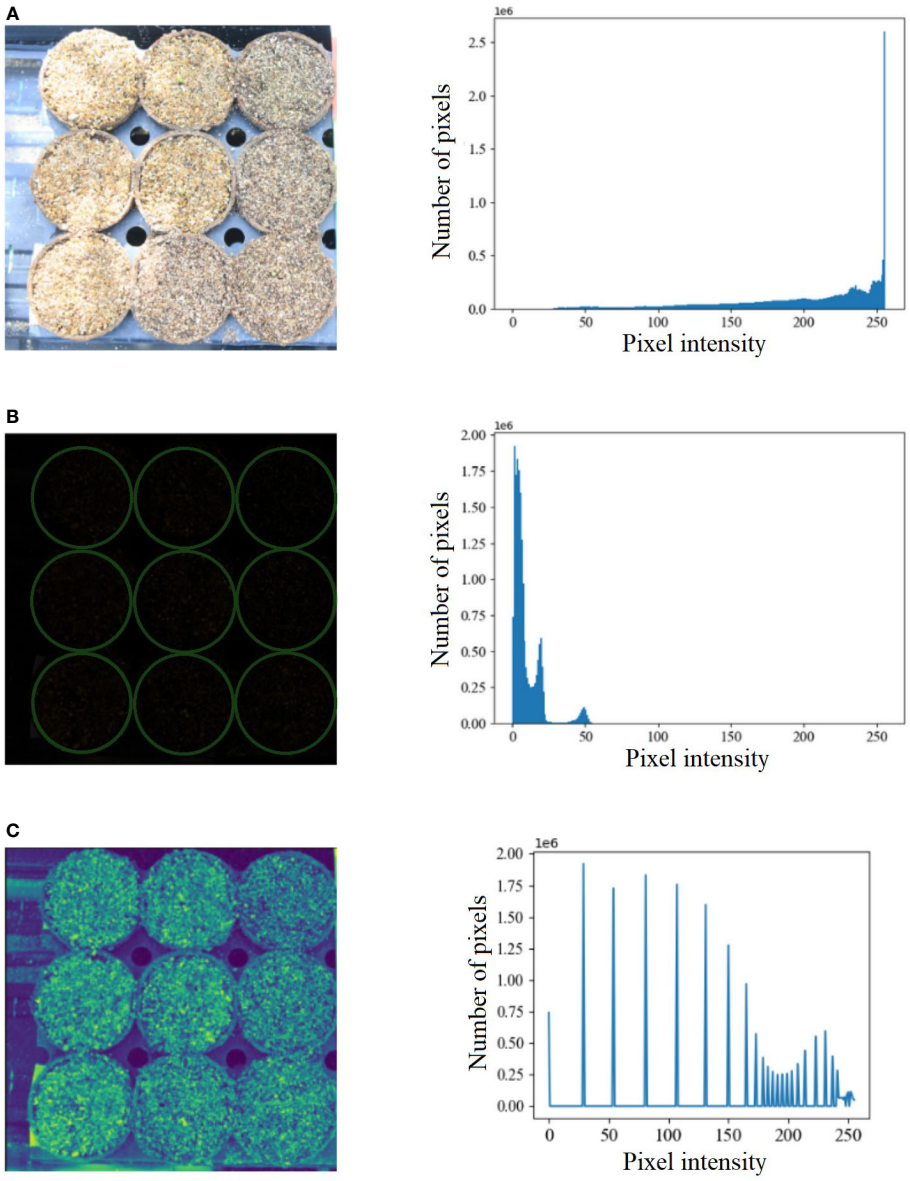


FIGURE 1
(A) An example of the brightest image from UNL-MED and its corresponding histogram; **(B)** an example of the darkest image from UNL-MED (the pots are marked in green circles) and its corresponding histogram; and **(C)** the darkest image after histogram equalization and its corresponding histogram.

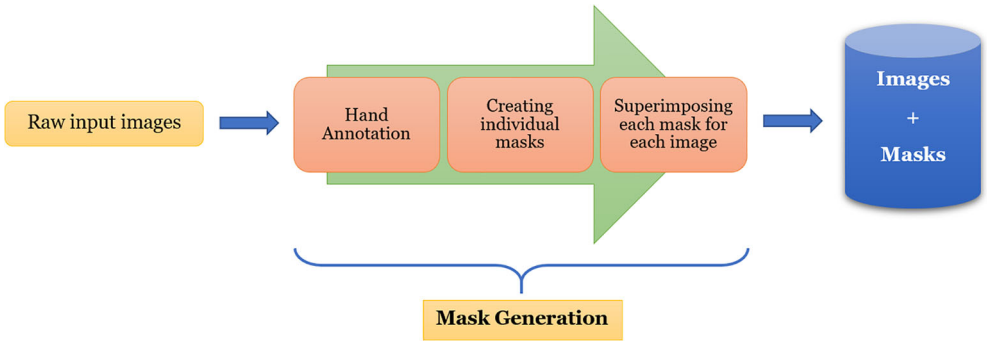


FIGURE 2
Illustration of mask generation process for UNL-MED.

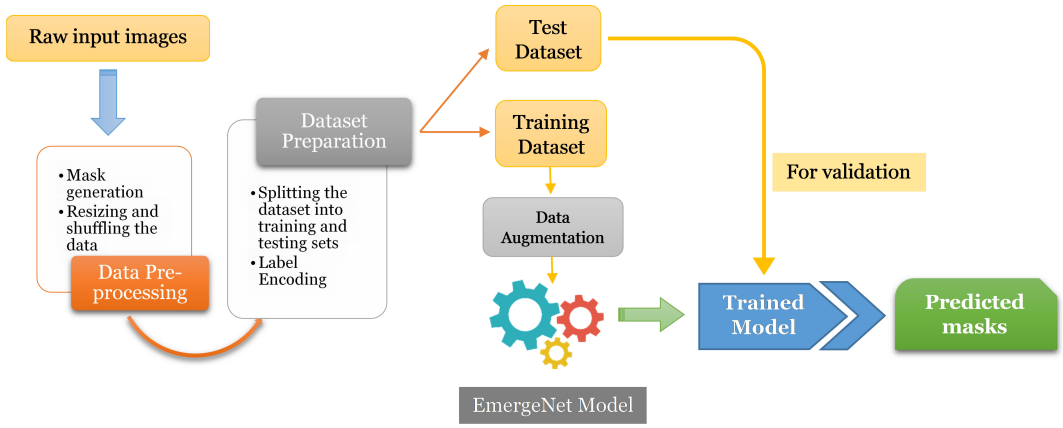


FIGURE 3
Flowchart of the proposed method.

behind UNet is to encode the image, passing it through a convolutional neural network as it gets downsampled, and then decode it back, or upsample it to obtain the segmentation mask. However, it is experimentally found that using a pre-trained model as its encoder and decoder, rather than using the standard UNet architecture, the performance of the model improves significantly (Lagree et al., 2021).

3.3.1.2 UNet with SEResNet backbone

A novel architectural unit called the squeeze-and-excitation (SE) block has been introduced in (Hu et al., 2018). It adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels at almost no computational cost. This is achieved by mapping the input to the feature maps for any given transformation. A detailed description of the structure of SE block and its operational characteristics are provided in (Hu et al., 2018). As an example, adding SE blocks to ResNet50 results in almost the same accuracy as ResNet101, but at a much lower computational complexity.

3.3.1.3 UNet with InceptionV3 backbone

The Inception architecture, unlike conventional convolutional networks, is a very complex, heavily engineered, deep neural network that uses filters of multiple sizes operating at the same level, rather than stacked convolutional layers. It enhances the utilization of available computational resources as well as improves performance significantly. The main idea of the Inception architecture is to find out how an optimal local sparse structure in a convolutional vision network can be approximated and covered by readily available dense components. InceptionV3 makes several improvements over earlier versions by including the following features: (a) *Label smoothing*, which is a regularization technique designed to tackle the problem of overfitting as well as overconfidence in deep neural networks; (b) *Factorizing convolution* to reduce the number of connections/parameters without decreasing the network efficiency; and (c) *Auxiliary classifier* which is used as a regularizer. InceptionV3, with its 42-layer-deep network, is computationally cheaper and much more efficient than other deep neural networks (Szegedy et al., 2016).

3.3.1.4 UNet with VGG19 backbone

The VGG model (Simonyan and Zisserman, 2015) derives inspiration from its predecessor, AlexNet, and is a much-improved version that uses deep convolutional neural layers to achieve better accuracy. VGG19 is the successor to the VGG16 model with 19 layers. VGG19 achieves better accuracy (Shu, 2019) than the VGG16 model as it can extract features better with its deep convoluted network. VGG19 has 16 convolutional layers with 3 FC layers and 5 pooling layers. Here, 2 of the 3 FC layers consist of 4096 channels each. The final FC layer originally had 1000 channels, followed by a SoftMax layer.

We have used the previously discussed three models as the backbone for EmergeNet, replacing the encoder part of the UNet with one of the models at a time. Owing to the symmetric structure of the UNet model, in the decoder or the expansion path, we programmatically upscale the corresponding model in a symmetric fashion to get the final output. For example, if we use VGG19 as the backbone, we are replacing the encoder part of the UNet with the VGG architecture, and in the expansion path, we are using the same VGG architecture to programmatically upscale it. These backbone models were previously trained on the significantly large well-known dataset called 'ImageNet', which consists of 3.2 million images (Deng et al., 2009). Thus, using these pre-trained weights allows us to benefit from transfer learning for improved accuracy and speed.

3.3.2 EmergeNet loss function

Instead of the traditional 'binary cross-entropy' loss (the negative average of the log of corrected predicted probabilities), EmergeNet uses a weighted sum of the two loss functions which are relevant to the task of segmentation. They are the Dice coefficient loss and focal loss. The motivation for using these two loss functions instead of cross-entropy loss is that these functions address some of the limitations of traditional cross-entropy loss. The statistical distributions of labels play a big role in training accuracy when using cross-entropy loss. The training becomes more difficult as the label distributions become more unbalanced. This is because cross-entropy loss is calculated as the average of per-pixel loss without knowing whether its adjacent pixels are boundaries or not. The Dice

coefficient loss and the focal loss, discussed in detail, address these disadvantages and therefore boost the performance of the model.

The Dice coefficient is a statistic used to gauge the similarity of two samples and was independently developed by Thorvald Sørensen and Lee Raymond Dice (Sørensen–Dice coefficient, 1948). It was brought to the computer vision community by (Milletari et al., 2016) for 3D medical image segmentation. The Dice loss is computed by

$$D = \frac{2*\sum_i^N p_i*g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (1)$$

where, p_i and g_i represent pairs of corresponding pixel values of prediction and ground-truth, respectively. The values of p_i and g_i are either 0 or 1 in boundary detection scenarios, therefore the denominator becomes the sum of the total boundary pixels of both prediction and ground-truth and the numerator becomes the sum of correctly predicted boundary pixels because the sum increments only when p_i and g_i match (both of value 1). Figure 4A shows the Venn diagram for the Dice loss.

The Dice similarity coefficient (DSC) is a measure of the overlap between two sets (see Figure 4A). In the task of boundary detection, the ground-truth boundary pixels and predicted boundary pixels can be viewed as two sets. By leveraging Dice loss, the two sets are trained to overlap gradually as training progresses. The denominator considers the total number of boundary pixels at a global scale, while the numerator considers the overlap between the two sets at a local scale. Therefore, DSC considers the loss information both locally and globally, making it a very effective loss metric for segmentation.

Focal loss, developed by (Lin et al., 2017), is a modified version of the Cross-Entropy (CE) loss. In the focal loss, the loss for correctly classified labels is scaled down so that the network focuses more on incorrect and low-confidence labels. In the task of segmenting a tiny foreground that relies on pixel-wise classification, a huge class imbalance occurs due to the presence of a considerably large background. Easily classified negatives comprise the majority of the loss and dominate the gradient. The focal loss is designed to address this issue by modifying the CE loss equation. The CE loss for binary classification is given as:

$$CE(p, y) = \begin{cases} -\log(p), & \text{if } y = 1 \\ -\log(1-p), & \text{otherwise} \end{cases} \quad (2)$$

where $y \in \pm 1$ specifies the ground-truth class and $p \in [0,1]$ is the model's estimated probability for the class with label $y=1$.

(Lin et al., 2017) proposed to reshape the loss function to down-weigh easy examples and therefore focus more on training on hard negatives. Mathematically, they proposed to add a modulating factor $(1-p_t)^\gamma$ to the CE loss, with tunable focusing parameter $\gamma \geq 0$. Focal Loss is therefore defined as:

$$FL(p_t) = -(1-p_t)^\gamma \log(p_t) \quad (3)$$

where p_t is given by the equation:

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases} \quad (4)$$

The focal loss is visualized for several values of $\gamma \in [0,5]$ in Figure 4B.

From Figure 4B, we note the following properties of the focal loss:

- When an example is misclassified and p_t is small, the modulating factor is near 1 and the loss is unaffected.
- As $p_t \to 1$, the factor goes to 0 and the loss for well-classified examples is down-weighted.
- The focusing parameter γ, smoothly adjusts the rate at which easy examples are down-weighted.

The EmergeNet Loss function, $loss_e$, is calculated by the weighted sum of Dice coefficient loss, i.e., $loss_d$ (defined in Eq 1), and focal loss, i.e., $loss_f$ (defined in Eq 3) as follows:

$$loss_e = loss_d + \alpha \times loss_f \quad (5)$$

where α is the tuning factor. For our experiment, it has been experimentally found that the optimal value of α is 1.

3.3.3 Performance-based weighted ensemble learning

Ensemble learning is a process by which multiple models are strategically generated and combined to solve a particular computational intelligence problem for improved performance. An ensemble model is typically constructed in two steps. First, a number of base learners are built either in parallel or in a sequence. Then, the base learners are combined using popular techniques like majority voting or weighted averaging. There are three main reasons

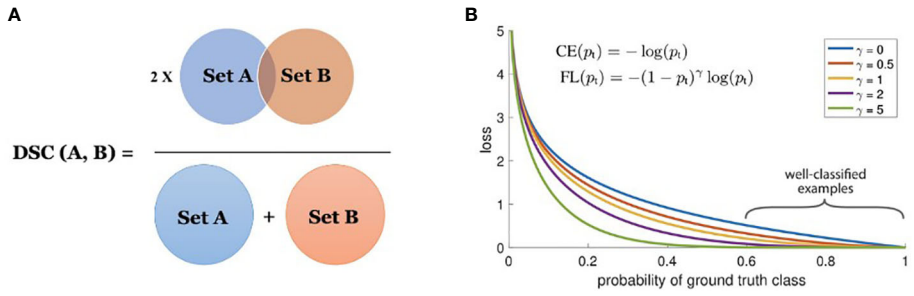


FIGURE 4
(A) Dice coefficient (set view); and **(B)** focal loss for $\gamma \in [0,5]$.

(Dietterich, 1997) why the generalization ability of an ensemble is usually much stronger than that of a single learner:

- The training data might not provide sufficient information for choosing a single best learner. For example, many base learners could perform equally well on the training dataset. Therefore, combining these learners might be a better choice.
- The search processes of the learning algorithms might be imperfect. For example, it might be difficult to achieve a unique best hypothesis, even if one exists, since the algorithms result in a sub-par hypothesis. This can be mitigated by the use of ensemble learning.
- The hypothesis space being searched for might not contain the true target function, while ensembles can give some good approximation.

Instead of using state-of-the-art ensembling techniques like bagging or boosting, EmergeNet introduces a novel weighted ensembling technique that aims to calculate the weights of the individual models based on their performances. These weights are then used to reward or penalize the models. Let IoU_i be the Intersection over Union (IoU) score of the ith model. We define a penalizing factor p_i as:

$$p_i = \frac{(100 - IoU_i)^2}{\sum (100 - IoU_i)^2} \tag{6}$$

The optimal weight, w_i is then calculated as:

$$w_i = \frac{\frac{1}{p_i^2}}{\sum \frac{1}{p_i^2}} \tag{7}$$

Finally, the ensembled weighted IoU of the EmergeNet (IoU_w) is computed as follows:

$$IoU_w = \sum IoU_i \cdot w_i \tag{8}$$

Figure 5 shows a compact view of the proposed EmergeNet architecture.

3.4 Evaluation metrics

The performance of our proposed EmergeNet model has been evaluated using three evaluation metrics, namely, F1-Score, Matthews Correlation Coefficient (MCC) (Matthews, 1975), and Intersection over Union (IoU) whereas the emergence time detection is evaluated using our proposed Emergence Time Accuracy (ETA). Accuracy (or Pixel Accuracy) is not a reliable metric for the task of segmenting tiny objects because this metric is strongly biased by classes that take a large portion of the image. Therefore, we have not used accuracy as a performance metric in this study. It is worth noting that a True Positive (TP) is an outcome where the model correctly predicts the positive class. Similarly, a True Negative (TN) is an outcome where the model correctly predicts the negative class. A False Positive (FP) is

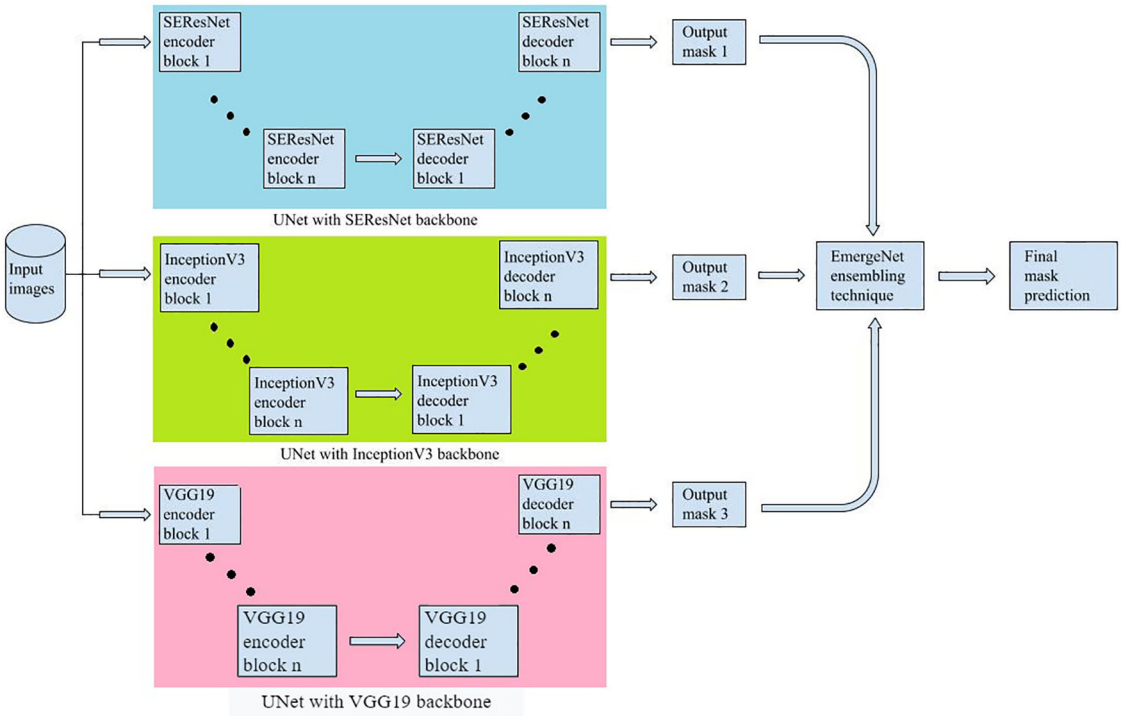


FIGURE 5
The proposed EmergeNet architecture.

an outcome where the model incorrectly predicts the positive class and a False Negative (FN) is an outcome where the model incorrectly predicts the negative class. These metrics are defined as follows:

- F1-Score is the Harmonic Mean between precision and recall. The range for F1-Score is [0, 1], with 0 being the worst and 1 being the best prediction. It is governed by the equation:

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{9}$$

- MCC is an improved metric which takes into account true and false positives and negatives and is generally regarded as a balanced measure that can be used even if the classes are of very different sizes. It has a range of -1 to 1 where -1 is a completely negative correlation between ground-truth and predicted value whereas +1 indicates a completely positive correlation between the ground-truth and predicted value.

$$\text{MCC} = \frac{\text{TP} \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{10}$$

- Intersection over Union (IoU) is a number from 0 to 1 that specifies the amount of overlap between the prediction and ground-truth.

$$IoU = \frac{Area\ \ of\ \ Overlap}{Area\ \ of\ \ Union} \tag{11}$$

The emergence time is defined as the timestamp of the image in which EmergeNet first detects the coleoptile(s).

Let $M=\{\alpha_1,\alpha_2,\ldots,\alpha_n\}$, where α_i denotes the image for a seeded pot obtained at timestamp t_i, n denotes the total number of images in the sequence, where $t_i < t_{i+1}$, $\forall\ 1 \leq i < n$. The emergence time for a pot is given by the first timestamp EmergeNet finds the coleoptile. Thus,

$$EmergenceTime(M) = j : EmergeNet(j) \neq \emptyset\ \ AND\ \ EmergeNet(i) = \emptyset, 1 \leq i < j \tag{12}$$

The emergence time accuracy (ETA) is determined by comparing the time computed from the results from EmergeNet with the ground-truth, obtained by careful manual inspection of the image sequence.

Given an image sequence, the detection of the emergence time is considered accurate if the time predicted based on the results of EmergeNet (Eq 12) matches the ground-truth, i.e.,

$$EmergenceTime(M) = GroundTruth(M), \tag{13}$$

where $GroundTruth(M)$ is the timestamp of the emergence determined manually.

ETA is given by the proportion of emergences accurately identified by EmergeNet. Thus,

$$ETA = \frac{Number\ \ of\ \ emergences\ \ correctly\ \ detected\ \ by\ \ EmergeNet \times 100\,\%}{Total\ \ number\ \ of\ \ emergences\ \ in\ \ the\ \ image\ \ sequences} \tag{14}$$

4 Experimental analysis

This section discusses the experimental setup, the benchmark dataset, the evaluation metrics used to evaluate the proposed method, and the results obtained from our experiments.

4.1 Experimental setup

The experimental analyses are performed using the Kaggle Notebook, a cloud computational environment that provides a free platform to run code in Python using dedicated GPUs. Kaggle Notebooks run in a remote computational environment and each Notebook editing session is provided with many resources. We used a GPU Kernel with Tesla P100 16 GB VRAM as GPU, with 13 GB RAM along with a 2-core of Intel Xeon as CPU. The training masks are generated using the VGG annotator tool. Python is featured with a plethora of useful packages, like, OpenCV, TensorFlow, Keras, Scikit-learn, etc., which are used to train the model and evaluate its performance. The number of images used for training was 260 (130 images and their corresponding masks). The execution time for training the EmergeNet was 1.5 hours. Compared to other deep neural networks, EmergeNet took less time to train as it benefits from transfer learning. We trained each model until the IoU curve for each of them reached saturation, and no further improvement was possible.

4.2 Results

We present our results in two different parts. First, we present a comparative study of EmergeNet and UNet in terms of performance. We have also compared the performance of all three individual backbone networks with EmergeNet. A more detailed analysis of the performance of EmergeNet under dark lighting conditions is also presented. In the second part, we analyzed the growth monitoring of maize coleoptiles as well as their emergence timing detection.

4.2.1 Comparative study

Figure 6 shows the comparison of masks generated by UNet and EmergeNet using a test image sequence from the UNL-MED. Figure 6A shows the masks generated by the standard UNet model and their corresponding ground-truth. Note that the generated masks do not accurately match with the ground-truth. Furthermore, the UNet model failed to detect the emergence of coleoptiles in several cases. This is generally the case with generic models which are not sophisticated enough. The IoU obtained by the standard UNet model for this test sample is 69%. Figure 6B shows the masks generated by the proposed EmergeNet model and their corresponding ground-truth.

In contrast to the UNet model's results, the masks generated using EmergeNet closely correspond to the ground-truth. EmergeNet neither incorrectly generated a mask (when there was no emergence), nor failed to produce a mask when there was a

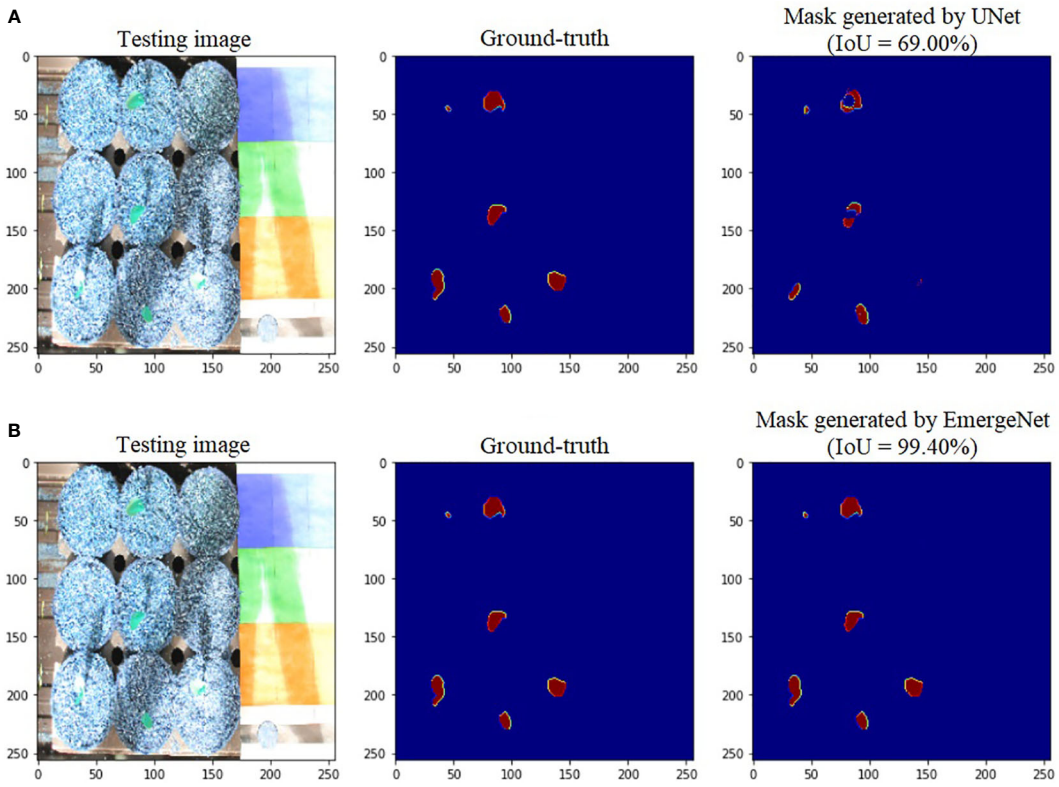


FIGURE 6
Illustration of segmentation performance on a test image from UNL-MED by **(A)** UNet and **(B)** EmergeNet.

coleoptile. The overall IoU of EmergeNet for this test sample is 99.40%, a significant improvement over UNet. Table 1 shows the comparative analysis of UNet and EmergeNet in terms of the three evaluation metrics, namely, F1-Score, MCC, and IoU, for all images of UNL-MED. It is evident from the table that EmergeNet significantly outperforms the UNet model.

While a confusion matrix (CM) is a powerful visualization technique to summarize the performance of a supervised classification task, it does not provide valuable insights into the model's performance for image segmentation since the data is highly imbalanced toward the background class. The normalized CMs for a random test image for the standard UNet and EmergeNet are shown in Figure 7. It is evident from the figure that the background is significantly larger than the maize coleoptile. A more accurate representation of the classifier's performance can be derived by overlaying the values of the confusion matrix on the coleoptile mask generated by the classifier. Figure 7A shows that majority of the mask generated by the standard UNet (shown in magenta) is incorrectly labeled as the coleoptile, i.e., false positive. Only a very small portion of the mask is accurately labeled (shown in cyan), denoting the true positives. Figure 7B shows the mask overlaid by the values of the corresponding confusion matrix for EmergeNet. It shows that a significant majority of the mask is correctly labeled (shown in cyan), and only a few pixels, mostly along the border, are false positives (shown in magenta). There are no false negatives or true negatives for EmergeNet. This demonstrates the efficacy of EmergeNet and its superiority over the standard UNet for this application.

TABLE 1 Results of comparison between UNet and EmergeNet on UNL-MED under all lighting conditions in terms of F1-Score, MCC, and IoU.

	F1-Score	MCC	IoU
UNet	0.6813	0.7040	0.7690
EmergeNet	0.9470	0.9894	0.9820

To further add credibility to the accuracy and robustness of the integrated network, i.e., EmergeNet, we performed a comparative study among the individual networks with EmergeNet. In each case, the masks generated by EmergeNet are better than that created by the individual networks. Figure 8 compares the masks generated by UNet with SEResNet as its backbone, and EmergeNet. The IoU of EmergeNet is higher by 2.21%. InceptionV3 is a very powerful network on its own, and therefore, the UNet structure with InceptionV3 as its backbone is expected to perform remarkably well. Such is the case as depicted in Figure 9, however, EmergeNet still beats the IoU score by 0.11% which is impressive considering the fact that it becomes exponentially more difficult to improve the results above a certain threshold value. Finally, EmergeNet beats UNet with VGG as its backbone by 0.54% in terms of IoU metrics as shown in Figure 10. Conclusively, it can be inferred that the integrated structure of EmergeNet plays a significant role in bringing the best of all the individual networks and performs better than all of them, thereby proving its worth as a segmentation model.

To demonstrate the efficacy of EmergeNet under extremely low light conditions, we conducted the same experimental analyses by

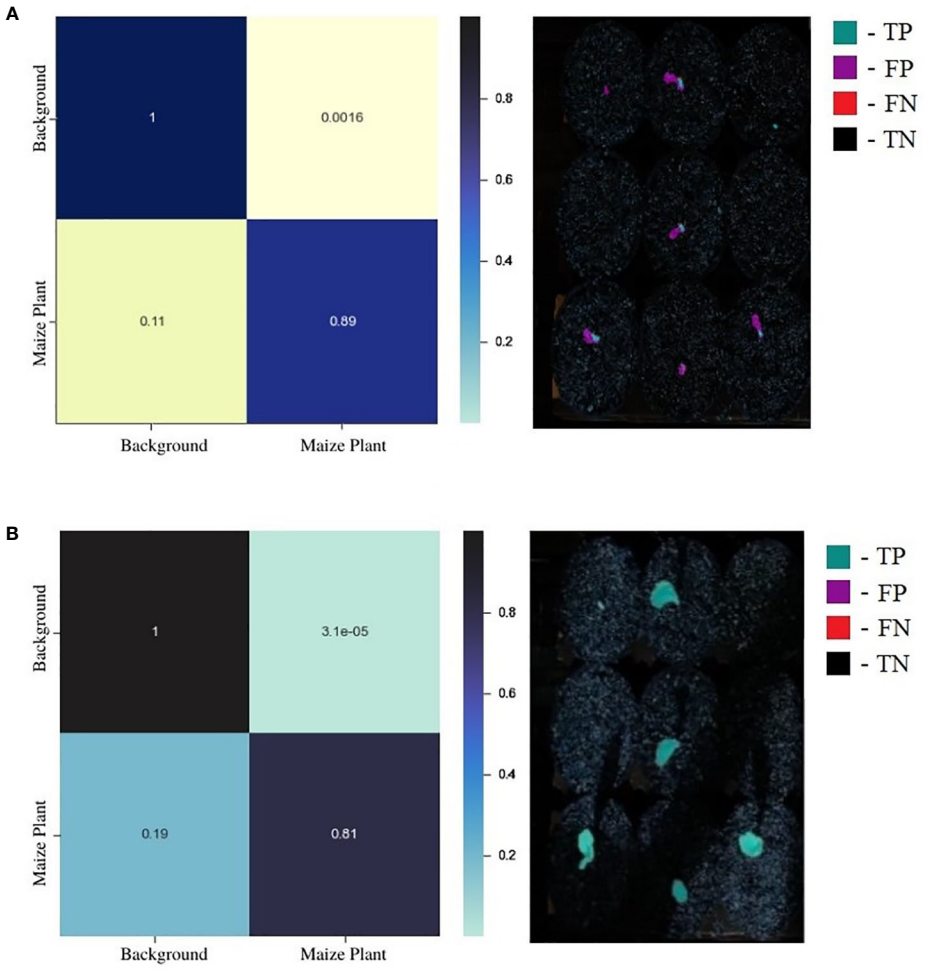


FIGURE 7
(A) Juxtaposition of confusion matrix (left) and its corresponding overlay mask from standard UNet (right). **(B)** Juxtaposition of confusion matrix (left) and its corresponding overlay mask from the EmergeNet (right).

considering the images of UNL-MED that were captured in a dark environment only. Table 2 summarizes the results of the comparison between the UNet and EmergeNet in dark conditions. Results show that EmergeNet significantly outperformed the standard UNet along all three evaluation metrics.

4.2.2 Growth monitoring

Thus, EmergeNet can efficiently monitor the growth of maize coleoptiles even at extremely low light conditions. We defined a new measure called the ETA to evaluate the accuracy of emergence in Section 4.4. Figure 11 displays a sequence of images that show the emergence and growth of coleoptiles computed by EmergeNet. It shows that EmergeNet detected the emergence of coleoptile with 100% ETA. Furthermore, it correctly tracks the growth of the coleoptile, as demonstrated by the increasing size of the generated masks in the time-lapse imagery. Note that the first two subfigures, i.e., Figures 11A, B, do not contain any masks because the emergence has not taken place yet. Out of nine maize seeds sown in the nine pots, only four of them emerged earlier, as shown by the tiny masks in Figure 11C. Subsequently two more coleoptiles emerged, one in Figure 11D and the other in Figure 11G. Three seeds failed to emerge in the experiment. After their emergence, all six coleoptiles are correctly tracked. Thus, EmergeNet accurately identifies the emergence events and tracks the growth of the coleoptiles over time.

We conducted an experiment to demonstrate the accuracy of the coleoptile size measured by the total number of constituent pixels of the generated masks by comparing them with the ground-truth. The result of the comparison is shown in Figure 12. The figure shows the coleoptile size of the generated mask (shown in blue) significantly overlaps with the coleoptile size of the ground-truth (shown in red). Thus, EmergeNet not only produces high-precision masks but also helps in the growth monitoring of coleoptiles with very high accuracy. Table 3 shows the Pearson correlation coefficient between the coleoptile size of the mask generated by EmergeNet and the ground-truth. Pearson correlation coefficient is a measure of the linear relationship between two variables. It has a value between -1 to 1, with a value of -1 denoting a total negative

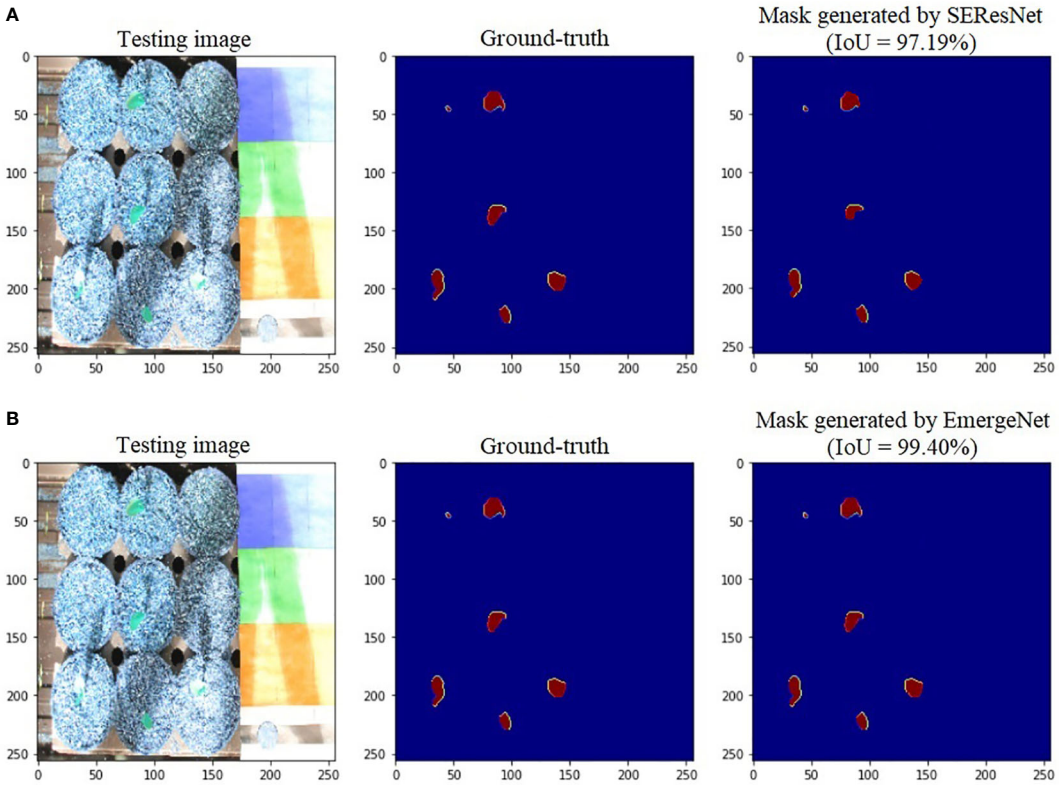


FIGURE 8
Illustration of segmentation performance on a test image from UNL-MED by **(A)** SEResNet and **(B)** EmergeNet.

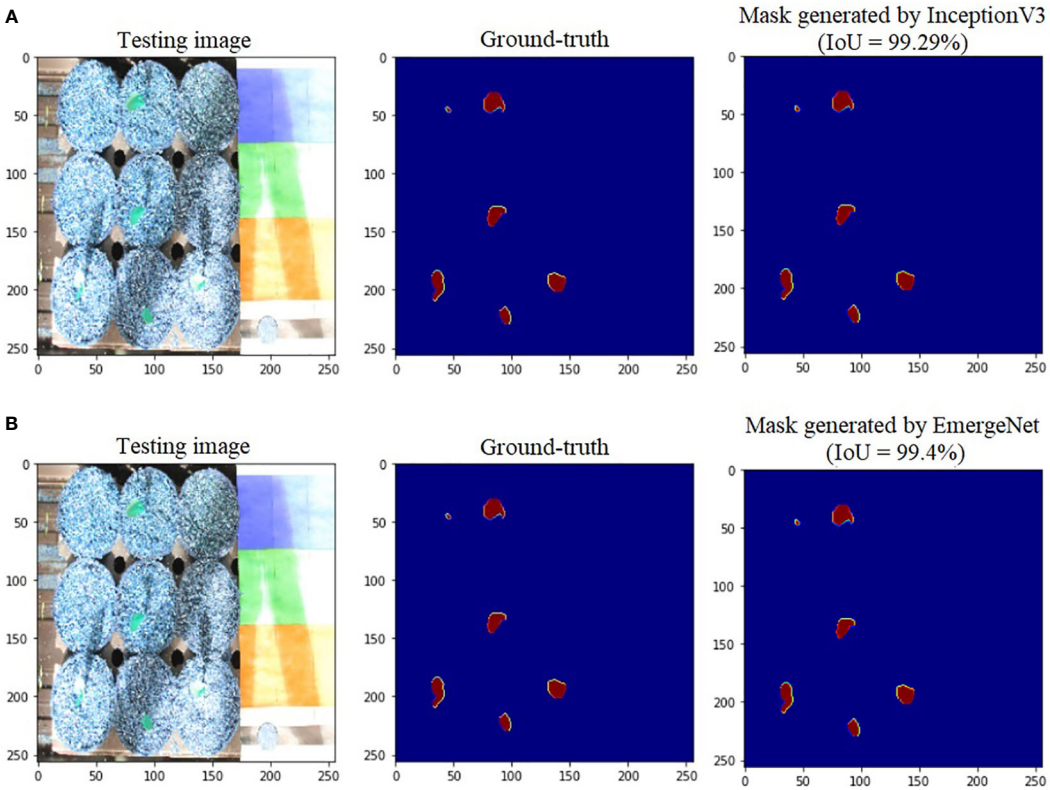


FIGURE 9
Illustration of segmentation performance on a test image from UNL-MED by **(A)** InceptionV3 and **(B)** EmergeNet.

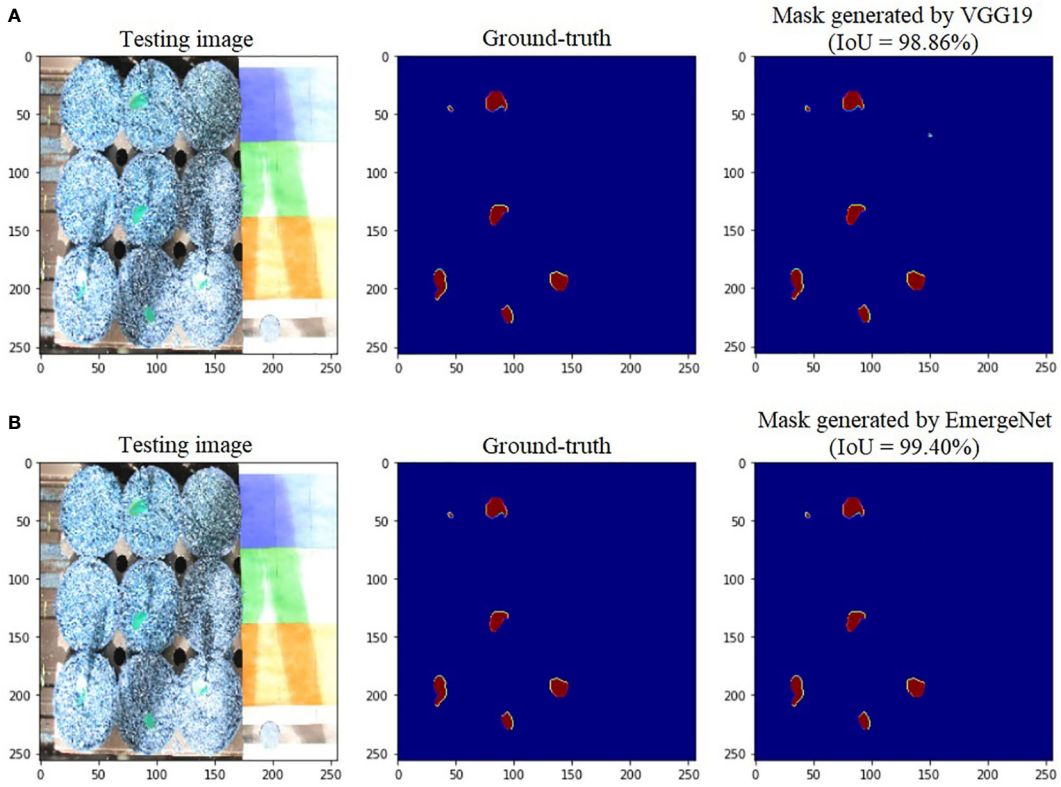


FIGURE 10
Illustration of segmentation performance on a test image from UNL-MED by **(A)** VGG19 and **(B)** EmergeNet.

linear correlation, 0 being no correlation, and + 1 denoting a total positive correlation. The table shows a high positive correlation between the ground-truth mask and the generated mask in terms of coleoptile size.

5 Discussion

The timing of germination is a paramount physiological factor for seed quality determination that encompasses a set of broad concerns, including vigor, dormancy mechanisms, pests, pathogens, genetic integrity, cost of establishment, field maintenance to prevent contamination with weeds or unwanted seed, and isolation distances to prevent cross-pollination Stiller et al. (2010). Thus, research attention for automated emergence timing determination based on computer vision and artificial intelligence techniques to replace tedious manual human labor is more crucial than ever. In this paper, we proposed an ensemble deep-learning based segmentation model based on the UNet architecture for coleoptile emergence time detection. The proposed EmergeNet outperforms the UNet by a significant margin as demonstrated by the experimental analyses in Section 4.2. The success of EmergeNet is attributed to successful base backbone architectures, customized loss function, and a novel penalizing factor in the ensemble technique. The accuracy of any image-based phenotypes depends on the accuracy of the underlying segmentation model Das Choudhury (2020). Thus, the ensemble segmentation model introduced in this paper has the potential to be extended to other automated phenotyping applications.

TABLE 2 Results of comparison between UNet and EmergeNet on UNL-MED only under dark environmental conditions in terms of F1-Score, MCC, and IoU.

	F1-Score	MCC	IoU
UNet	0.6065	0.6000	0.7774
EmergeNet	0.9709	0.9709	0.9735

The proposed EmergeNet model significantly outperforms the widely used UNet architecture. The IoU metric for mask generation for EmergeNet is 99.4%; in comparison, the standard UNet has an IoU of 69% (see Table 1). By combining the three UNet architectures with powerful backbones, along with the proposed custom loss function and a novel ensemble technique, EmergeNet reduces the number of false positives (see Figure 7B). One of the key contributions of this model is its success in extremely low light, as evident from Table 2. Overall, EmergeNet is able to detect the emergence timing of the maize coleoptile with 100% accuracy. Furthermore, EmergeNet can accurately monitor the growth of the coleoptiles over time, as demonstrated by a very high correlation (0.999) between the generated masks and the ground-truth.

6 Conclusion

The timing of important events in a plant's life, for instance, germination, the emergence of a new leaf, flowering, fruiting, and onset of senescence, is crucial in the understanding of the overall

10.3389/fpls.2023.1084778

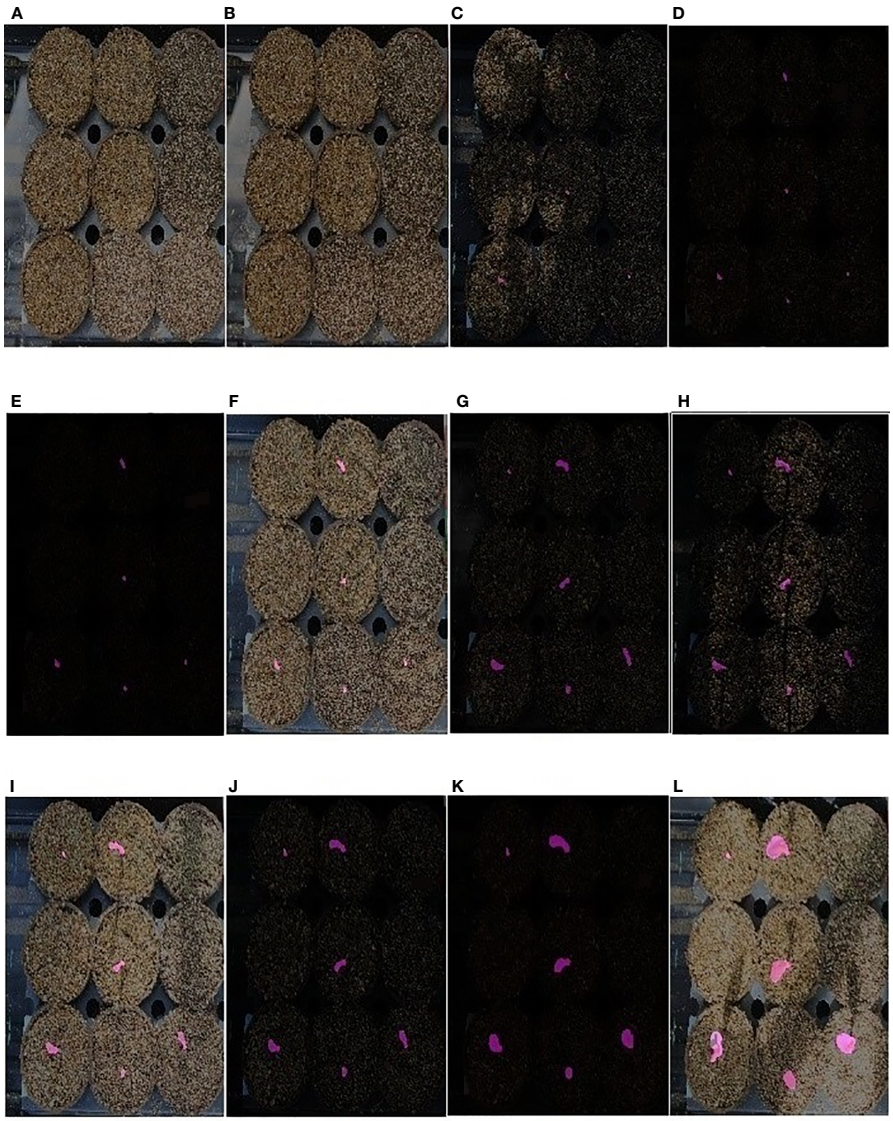


FIGURE 11
(A–L) Illustration of emergence timing detection and growth monitoring of a maize coleoptile in a time-lapse test image sequence of UNL-MED.

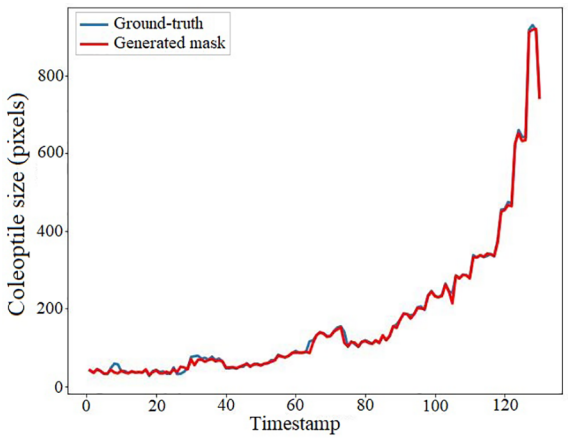


FIGURE 12
Coleoptile size measured by the total number of constituent pixels of the ground-truth (shown in blue) and the generated mask (shown in red) as a function of time.

plant's vigor, which is likely to vary with the interaction between genotype and environment, and are referred to as event-based phenotypes. This paper introduces a novel deep-learning model called EmergeNet to detect the timing of the emergence of a maize seedling and track its growth over a time-lapse video sequence. EmergeNet is based on an ensemble model that integrates SEResNet18, InceptionV3, and VGG19, such that it overcomes the challenge of detecting a tiny living object and tracks its changes in shape and appearance in the presence of cluttered soil background and extreme variation of illuminations. Furthermore, the paper introduces a benchmark dataset

TABLE 3 Pearson correlation table between ground-truth and generated mask.

	Ground-truth	Generated mask
Ground-truth	1.000000	0.999412
Generated mask	0.999412	1.000000

called UNL-MED. Experimental evaluation on UNL-MED shows the capability of EmergeNet to detect the timing of emergence with 100% accuracy as compared with human-perceived ground-truth. It is also experimentally demonstrated that EmergeNet significantly outperforms its base model UNet in the task of segmentation. EmergeNet incorporates three pre-trained networks including all their weights, and hence, it requires high-end computing power for efficient training. Additionally, EmergeNet is trained on only one type of plant constrained by a set of external environmental conditions. Future work will consider the detection of multiple coleoptiles with or without the presence of weeds in the same pot.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://plantvision.unl.edu/dataset.

Author contributions

AD developed and implemented the algorithm, conducted experimental analysis, and contributed to manuscript writing. SD conceived the idea, led the dataset design, conducted experimental analysis, led the manuscript writing, and supervised the research. AS, AKD, and TA critically reviewed the manuscript and provided constructive feedback throughout the process. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by Agricultural Genome to Phenome Initiative Seed Grant [grant no. 2021-70412-35233] from the USDA National Institute of Food and Agriculture.

Acknowledgments

Authors are thankful to Dr. Vincent Stoerger, the Plant Phenomics Operations Manager at the University of Nebraska-Lincoln, USA, for his support in setting up experiments to create the dataset used in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agarwal, B. (2017). Detection of Plant Emergence Based on Spatio Temporal Image Sequence Analysis. Master's thesis (The University of Nebraska-Lincoln).

Aggarwal, J. K., and Cai, Q. (1999). Human motion analysis: A review. *Comput. Vision image understanding* 73, 428–440. doi: 10.1006/cviu.1998.0744

Arulmurugan, A., Hema Rajini, N., and Pradeep, G. (2021). Ensemble of deep learning based segmentation with classification model for skin lesion diagnosis using dermoscopic images. *J. Comput. Theor. Nanoscience* 18, 710–721. doi: 10.1166/jctn.2021.9667

Bashyam, S., Das Choudhury, S., Samal, A., and Awada, T. (2021). Visual growth tracking for automated leaf stage monitoring based on image sequence analysis. *Remote Sens.* 13. doi: 10.3390/rs13050961

Das Choudhury, S. (2020). "Segmentation techniques and challenges in plant phenotyping," in *Intelligent image analysis for plant phenotyping*. Eds. A. Samal and S. Das Choudhury (Boca Raton, Florida: CRC Press, Taylor & Francis Group), 69–91.

Das Choudhury, S., Bashyam, S., Qiu, Y., Samal, A., and Awada, T. (2018). Holistic and component plant phenotyping using temporal image sequence. *Plant Methods* 14. doi: 10.1186/s13007-018-0303-x

Das Choudhury, S., and Samal, A. (2020). "Structural high-throughput plant phenotyping based on image sequence analysis," in *Intelligent image analysis for plant phenotyping*. Eds. A. Samal and S. Das Choudhury (Boca Raton, Florida: CRC Press, Taylor & Francis Group), 93–117.

Das Choudhury, S., Samal, A., and Awada, T. (2019). Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00508

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition (Ieee)* (Miami, FL, USA), 248–255.

Dhaka, V. S., Meena, S. V., Rani, G., Sinwar, D., Ijaz, M. F., and Woźniak, M. (2021). A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. *Sensors* 21, 4749. doi: 10.3390/s21144749

Dietterich, T. G. (1997). Machine-learning research. *AI magazine* 18, 97–97. doi: 10.1609/aimag.v18i4.1324

Dollár, P., Wojek, C., Schiele, B., and Perona, P. (2009). "Pedestrian detection: A benchmark," in *2009 IEEE conference on computer vision and pattern recognition (IEEE)* (Miami, FL, USA), 304–311.

Dutta, A., Gupta, A., and Zissermann, A. (2016) *VGG image annotator (VIA)*. Available at: http://www.robots.ox.ac.uk/vgg/software/via/.

Golla, A.-K., Bauer, D. F., Schmidt, R., Russ, T., Nörenberg, D., Chung, K., et al. (2020). Convolutional neural network ensemble segmentation with ratio-based sampling for the arteries and veins in abdominal ct scans. *IEEE Trans. Biomed. Eng.* 68, 1518–1526. doi: 10.1109/TBME.2020.3042640

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Salt Lake City, UT, USA), 7132–7141.

Kim, Y.-W., Byun, Y.-C., and Krishna, A. V. (2021). Portrait segmentation using ensemble of heterogeneous deep-learning models. *Entropy* 23, 197. doi: 10.3390/e23020197

Lagree, A., Mohebpour, M., Meti, N., Saednia, K., Lu, F.-I., Slodkowska, E., et al. (2021). A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-021-87496-1

Li, Y., Fan, X., Mitra, N. J., Chamovitz, D., Cohen-Or, D., and Chen, B. (2013). Analyzing growing plants from 4d point cloud data. *ACM Trans. Graphics* 32. doi: 10.1145/2508363.2508368.Yangyan2013

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision* (Venice, Italy), 2980–2988. lin2017focal.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta (BBA)-Protein Structure* 405, 442–451. matthews1975comparison. doi: 10.1016/0005-2795(75)90109-9

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV) (IEEE)* (Stanford, CA, USA), 565–571. milletari2016v.

Min, W., Fan, M., Guo, X., and Han, Q. (2018). A new approach to track multiple vehicles with the combination of robust detection and two classifiers. *IEEE Trans. Intelligent Transportation Syst.* 19, 174–186. doi: 10.1109/TITS.2017.2756989Min2018

Sørensen–Dice coefficient (1948) *Sørensen–dice coefficient — Wikipedia, the free encyclopedia* (Accessed 23-July-2021). enwiki:1023592604.

Scharr, H., Bruns, B., Fischbach, A., Roussel, J., Scholtes, L., and Stein, J. v. (2020). "Germination detection of seedlings in soil: A system, dataset and challenge," in *Computer vision - ECCV 2020 workshops*. Eds. A. Bartoli and A. Fusiello (Springer International Publishing), 360–374. Hanno2020.

Shu, M. (2019). *Deep learning for image classification on very small datasets using transfer learning shu* (MS thesis, Iowa State University, USA).

Simonyan, K., and Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition* (San Diego, CA, USA:International Conference on Learning Representations), arXiv preprint arXiv:1409.1556 vgg.

Stiller, B., Bocek, T., Hecht, F., Machado, G., Racz, P., and Waldburger, M. (2010). *Mobile systems IV* (Zurich, Switzerland:Tech. rep., University of Zurich, Department of Informatics instance), 1290.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, Nevada, USA), 2818–2826. [Dataset].

Wang, Z., Walsh, K., and Koirala, A. (2019). Mango fruit load estimation using a video based mangoyolo–kalman filter–hungarian algorithm method. *Sensors* 19, 2742. Zhenglin2019. doi: 10.3390/s19122742

Xing, J., Ai, H., Liu, L., and Lao, S. (2011). Multiple player tracking in sports video: A dual-mode two-way bayesian inference approach with progressive observation modeling. *IEEE Trans. Image Process.* 20, 1652–1667. doi: 10.1109/TIP.2010.2102045. Xing2011.

Xu, Y., Zhou, X., Chen, S., and Li, F. (2019). Deep learning for multiple object tracking: A survey. *IET Comput. Vision* 13, 355–368. xu2019deep. doi: 10.1049/iet-cvi.2018.5598

Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *ACM computing surveys (CSUR* 38, 13–es. yilmaz2006object. doi: 10.1145/1177352.1177355

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 February 2023
DOI 10.3389/fpls.2023.1120182

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service, United States Department of Agriculture (USDA), United States

REVIEWED BY
Jing Zhou,
University of Wisconsin-Madison, United States
Frantisek Spoutil,
Institute of Molecular Genetics (ASCR), Czechia

*CORRESPONDENCE
Evangeline Corcoran
ecorcoran@turing.ac.uk

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 09 December 2022
ACCEPTED 13 February 2023
PUBLISHED 24 February 2023

CITATION
Corcoran E, Siles L, Kurup S and Ahnert S (2023) Automated extraction of pod phenotype data from micro-computed tomography.
Front. Plant Sci. 14:1120182.
doi: 10.3389/fpls.2023.1120182

COPYRIGHT
© 2023 Corcoran, Siles, Kurup and Ahnert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automated extraction of pod phenotype data from micro-computed tomography

Evangeline Corcoran[1]*, Laura Siles[2], Smita Kurup[2] and Sebastian Ahnert[1,3]

[1]Environment and Sustainability Theme, AI for Science and Government Programme, The Alan Turing Institute, London, United Kingdom, [2]Department of Plant Sciences for the Bioeconomy, Rothamsted Research, Harpenden, United Kingdom, [3]Department of Chemical Engineering and Biotechnology, School of Technology, University of Cambridge, Cambridge, United Kingdom

Introduction: Plant image datasets have the potential to greatly improve our understanding of the phenotypic response of plants to environmental and genetic factors. However, manual data extraction from such datasets are known to be time-consuming and resource intensive. Therefore, the development of efficient and reliable machine learning methods for extracting phenotype data from plant imagery is crucial.

Methods: In this paper, a current gold standard computed vision method for detecting and segmenting objects in three-dimensional imagery (StartDist-3D) is applied to X-ray micro-computed tomography scans of oilseed rape (*Brassica napus*) mature pods.

Results: With a relatively minimal training effort, this fine-tuned StarDist-3D model accurately detected (Validation F1-score = 96.3%,Testing F1-score = 99.3%) and predicted the shape (mean matched score = 90%) of seeds.

Discussion: This method then allowed rapid extraction of data on the number, size, shape, seed spacing and seed location in specific valves that can be integrated into models of plant development or crop yield. Additionally, the fine-tuned StarDist-3D provides an efficient way to create a dataset of segmented images of individual seeds that could be used to further explore the factors affecting seed development, abortion and maturation synchrony within the pod. There is also potential for the fine-tuned Stardist-3D method to be applied to imagery of seeds from other plant species, as well as imagery of similarly shaped plant structures such as beans or wheat grains, provided the structures targeted for detection and segmentation can be described as star-convex polygons.

KEYWORDS

phenotyping, plant development, machine learning, computer vision, micro-compute tomography

1 Introduction

The study of plant traits, such as plant architecture, growth, development, physiological or biochemical profiles is known as plant phenotyping. Identifying connections between plant genotype and phenotype is essential to advance our understanding of underlying developmental mechanisms in plant biology. With the recent rapid progression in functional genomics due to advances in high throughput sequencing, quantitative analyses of plant traits are of increasing relevance as it could allow for the links between genotype and phenotype to be explored in greater depth. For example, studies such as Genome-Wide Association Studies (GWAS) involve testing genetic variants across genotypes of a population to identify genotype-phenotype associations and provide essential information for plant breeding (Brachi et al., 2011; Alseekh et al., 2021). By obtaining insights on how genetics and environmental pressures lead to different phenotypic response in plants, more suitable and sustainable crops can be selected for growth in specific environments, as by identifying the genetic basis of phenotypic variation a better understanding of the factors driving plant adaptation and stress tolerance could be achieved.

The need to better characterize plant developmental growth stages and monitor traits that affect yield has led to an increased demand for and collection of high-throughput, high resolution plant image datasets (Costa et al., 2019). Analysis of such image datasets could allow for a more detailed understanding of dynamic developmental changes and for phenotypic traits to be measured in a non-destructive manner in comparison to current commonly used manual phenotyping methods. However, manual analysis of these image datasets can also be time-consuming, inconsistent, and requires expert observers. Therefore, developing reliable and efficient methods for automated extraction of phenotype data from plant images is crucial.

Image analysis pipelines for easy phenotyping have recently become more widely available such as those for measuring leaf area, leaf growth and root traits (Bours et al., 2012; Easlon & Bloom, 2014; Seethepalli et al., 2021). Although image acquisition is relatively straightforward, image analysis is plagued by a number of bottlenecks. In most cases, image thresholding and data extraction is still laborious and requires manual input. There also is a lack of consistency with regards to image acquisition between different days or between laboratories, which hampers the reliable extraction of phenotypic traits. Moreover, the automated images acquired in the agricultural sciences are driven by specific biological hypotheses, and the downstream pipelines typically are purpose-built and not compatible to other research areas, and often not free or easy to use.

In recent years, the study of plant organs and tissue development has been focused on the use of confocal microscopy, where 2D and 3D information is obtained by optical sectioning and the use of fluorescent markers. However, this is limited by the thickness of the sample being studied and the availability of suitable markers. A very valuable non-invasive and cost-effective 3D imaging technique for detecting and quantifying internal structures in a non-destructive manner without the necessity of using stains is X-ray micro-computed tomography (μCT), which is based on differential X-ray attenuation by biological materials. μCT scanners were developed mainly for medical purposes, and are not widely used in plant sciences (Dhondt et al., 2010; Pajor et al., 2013; Piovesan et al., 2021). To date, there has been limited application of this method to visualize above-ground plant structures because of the low attenuation density that these tissues present, resulting in images with low contrast (Pajor et al., 2013). However, recent improvements in scan resolution, quality and scan speed of current state-of-the-art μCT scanners present an opportunity to analyse these above-ground plant structures without the necessity of fixing or staining them. The μCT scanner has recently been used for the analysis of different plant tissues and organs, such as seeds, fruits, rice and wheat spikes flowers and leaves traits (Rousseau et al., 2015; Hughes et al., 2017; Tracy et al., 2017; Mathers et al., 2018; Schneider et al., 2018; Xiong et al., 2019; Gargiulo et al., 2020; Hu et al., 2020; Kunishima et al., 2020; Liu et al., 2020; Narisetti et al., 2020). These advances make this a promising technique to study complex plant traits, such as the internal structure of opaque mature pods without requiring destructive dissection methods. The resulting images have a higher resolution than those generated using other techniques such as light boxes or a light sheet confocal.

Although μCT scanning is a valuable tool for obtaining high-resolution images, advanced computational skills are required to develop automated data extraction pipelines from these images. This study aims to improve our understanding of seed biology and its related traits in the *Brassica napus* crop. For this purpose, images of mature seed pods were acquired, and data relating to the seed number per pod (SNPP) and seed area, as well as pod length were semi-automatically extracted. Although counting SNPP manually is quite easy, it is time consuming, and further data must be manually processed. When obtaining these data for GWAS studies with 100 individuals in a population, several hundreds of images need to be processed, resulting in an arduous and non-straightforward task. Moreover, more specific and biologically important information such as the position of the seeds in different pod valves and their relative spacing, is difficult to obtain. Therefore, we are interested in applying machine learning methods to assess whether this would allow us to generate a straightforward automated pipeline with minimal pre-processing for data extraction of phenotypic measurements from 3D μCT pod image data, including the number, size and shape of seeds, as well as their spatial arrangement relative to each other and to other pod structures.

In order to automatically extract valuable phenotypic measurements from 3D μCT pod image data, the first step required is to locate all individual seeds within a 3D volume. In machine learning, this is referred to as an object detection task, and can typically be achieved using models trained to recognize the target object (in this case *B. napus* seeds) and output the centre-point and/or a bounding box for each detection (Weigert et al., 2020). To extract data on seed size and shape, each pixel in the 3D volume needs to be labelled as either seed or background. Machine learning models designed to perform semantic segmentation have been shown to be able to achieve this for 3D volume data with high accuracy, but do not discern between individual objects meaning they cannot provide information on seed number and location (Alalwan et al., 2021; Kar et al., 2021;

Sodjinou et al., 2022). Therefore, an instance segmentation method that allows for both detection of multiple distinct objects, and that outputs both the number and location of seeds as well as a separate labelled mask capturing the shape and size of each seed is needed to achieve the goal of automatically extracting phenotype data from 3D µCT pod images (Lin et al., 2021; Wang et al., 2022).

Many automated instance segmentation methods have been proposed to process the increasingly large 3D volume datasets modern imaging instruments such as µCT scanners and microscopes are capable of producing (Meijering, 2012). These include non-machine learning approaches such as methods watershed transform-based morphological methods (Beucher & Meyer, 1993; Lotufo et al., 2002; Cheng & Rajapakse, 2009), graph-cut based optimization (Boykov & Funka-Lea, 2006), and thresholding or pixel-grouping using connected component analysis (Majanga & Viriri, 2021), as well as recent methods based on deep learning have been demonstrated to significantly improve the accuracy of instance segmentation predictions for images of biological specimens (Van Valen et al., 2016; He et al., 2017; Xie et al., 2018). These methods can be broadly sorted into two categories; methods in which semantic segmentation is performed first and pixels are then grouped into distinct objects (Çicek et al., 2016; Caicedo et al., 2019), and methods in which bounding boxes for individual objects are first predicted and then semantic segmentation is performed for each detected object (He et al., 2017; Xu et al., 2018; Zhao et al., 2018). However, despite the increased performance demonstrated by these deep learning methods in comparison to thresholding, watershed, and graph-cut optimization methods, they often still produce inaccurate results when used to predict the location and segment the individual shape of densely-packed objects, similar to the close positioning of seeds within the *B. napus* pod µCT image dataset (Schmidt et al., 2018).

StarDist-3D is an automated object detection and segmentation approach that was recently used to identify and examine the size, shape and spatial arrangement of individual cell nuclei in volumetric (3D) fluorescence microscopy images. It exhibits a high degree of accuracy in terms of both the predicted counts and shape of cell nuclei compared to other contemporary approaches such as U-Net and IFT-Watershed (Lotufo et al., 2002; Çicek et al., 2016; Schmidt et al., 2018; Weigert et al., 2020). The method uses a neural network to predict whether each pixel in a 3D volume is part of an individual target object, and to predict the distance to the object boundary using along several radial directions, defined based on spherical Fibonnaci lattice (Weigert et al., 2020). Thus, the shape of detected objects is predicted as a star-convex polygon, with increasing detail in the fluctuations along the surface of the object rendered when a higher number of Fibonnaci rays is used, allowing the anisotropy, or variation in shape of the target objects along different axial direction, of predicted objects to be more accurately reconstructed (Weigert et al., 2020).

It was determined that it would be potentially advantageous to apply the StarDist-3D approach to 3D µCT images of *B. napus* pods as the method was demonstrated to yield high accuracy in terms of both detection and segmentation for objects in close proximity to each other, as *B. napus* seeds are often observed to be tightly packed within pods (Weigert et al., 2020). Another potential advantage of the method was that it incorporates a computationally efficient non-maximum suppression (*nms*) process that reduces the likelihood of detecting the same object multiple times by suppressing detections with low confidence where the boundaries of detections with high confidence overlap (Schmidt et al., 2018; Weigert et al., 2020). Additionally, the StarDist-3D approach requires a relatively small amount of training data as it has been pre-trained to detect and segment a generalized variety of star-convex polygonal shapes. It is capable of quickly processing typically large 3D images, and the model outputs can easily be passed to state-of-the-art open-source toolkits for image analysis to extract specific data on the location, spatial arrangement, and 3D shape of predicted seeds (van der Walt et al., 2014; Gostick et al., 2019). Although the StarDist-3D has previously only been applied to cell microscopy images, it was posited that the method could be applied to any 3D volumetric image dataset regardless of sensor type or scale provided the target objects could be described as star-convex polyhedral. As *B. napus* seeds tend to be rounded or oval in shape, with slight variations or asymmetry along different axial directions, it was likely that the seeds could be appropriately described as star-convex polyhedra. Therefore, in this manuscript we explore the accuracy of detection and segmentation of a StarDist-3D model fine-tuned on 3D µCT images of *B. napus* pods, along with investigation of extraction of data on seed size, shape and spatial arrangement from the model outputs which could provide important biological information to improve models pertaining to plant development and crop yield.

2 Materials and methods

2.1 Plant growth conditions and data collection

A *B. napus* diversity set population with ninety-six genotypes was grown as in Siles et al. (2021). The seeds were germinated in P24 trays with John Innes Cereal Mix and once they presented four true leaves, they were transferred to a vernalization room with an 8 h photoperiod at 4°C day/night for 8 weeks. Each plant was re-potted in a 2 L pot in John Innes Cereal Mix. Each genotype had five biological replicates and once out of vernalization, all plants were grown in two glasshouse compartments in long-day conditions (16 h photoperiod) at 18°C day/15° night (600w SON-T, high pressure sodium lighting) at a density of 12 pots per m^2. Once the plants were fully dry and mature, the first five dry pods on the main inflorescence were ignored, and the next three developed pods were collected for scanning. To avoid pod shattering the pods were sprayed with Prism Clear Glaze (Loxley Arts, Sheffield, UK).

For each genotype, three fully dried pods were placed in plastic holders (34mm x 110mm) and packing peanuts were used to keep the samples in place while scanning. The pedicel was cut with a scalpel before placing the pods into the plastic holders. If the pods were too tall to fit in the holders, they were cut into two pieces and were separately scanned. Twelve holders were loaded into the

sample changing carousel of a µCT100 scanner (Scanco Medical, Switzerland). This scanner has a cone beam X-ray source with power ranging from 20 to 100 kVp (pre-set and calibrated for 45, 55, 70, 90 kVp) and a detector consisting of 3072 × 400 elements (48 µm pitch) and a maximum resolution of 1.25 µm. Pods were scanned with the X-ray power set at 45 kVp, 200 µA, 9W, with an integration time of 200 ms.

2.2 Image dataset description

Images were retrieved from the proprietary Scanco microCT file type format (.ISQ), which contained single-pixel width two-dimensional (2D) trans-axial projections, or 'slices', that together formed stacks depicting an entire pod as three-dimensional (3D) volumes. Thirty-two distinct 3D volumes were included in the experiment dataset, each containing a single entire *B. napus* pod. All 2D trans-axial (XY) slices were 512 × 512 pixels, therefore the height and width of all 3D volumes was also 512 pixels. Individual 3D volumes varied in length from 505 to 1397 slices, with a total of 29,871 slices in the experimental dataset. The total dataset contained 471 seeds.

The total dataset was split into a model training and validation dataset comprised of 13 3D volumes, 12,475 2D slices and 262 seeds and a model testing dataset containing comprised of 19 3D volumes, 17,396 2D slices and 209 seeds. This split was decided upon due to the uneven number of seeds in each seed pod, with the training and validation dataset containing 262(56%) of seeds and the testing dataset containing 209(44%) of seed. Another factor impacting the split of data was that intact 3D volumes of entire seedpods needed to be used for testing, to demonstrate that reliable seed detection and segmentation could be achieved on the original imagery without any pre-processing. Conversely, Weigert et al. (2020) demonstrated that more accurate results could be obtained in a computationally efficient manner by training a StarDist-3D on smaller sub-volumes of the original 3D volume data containing objects of interest, in this case sub-volumes containing at least one entire seed. The 3D volumes in the model training and validation dataset were therefore comprised of 138 small sub-volumes of stacked 2D slices containing a single seed, or multiple seeds in instances where seeds occupied some of the same 2D slices. These sub-volumes ranged in size between 24 to 84 2D slices depending on the size of the single seed or multiple overlapping seeds contained within. This sub-division was carried out in order to ensure a mixture of seeds from different seed pods could be used for model training and validation. 117 sub-volumes containing 220 seeds were randomly sorted into the final 'training' dataset, and 21 volumes containing 42 seeds were sorted into the final 'validation' dataset.

2.3 Image pre-processing and annotation

All 3D volumes contained in the experimental dataset were batch converted from their original.ISQ format into.TIF stacks using BoneJ plugin (Domander et al., 2021) for Fiji ImageJ software version 2.9.0 (Schindelin et al., 2012). All 262 seeds contained within the 138 3D sub-volumes comprising the 'training' and 'validation' datasets were then manually annotated using Fiji and the Labkit plugin (Schindelin et al., 2012; Arzt et al., 2022). Sub-volumes were converted from XYZ format to XYT timeseries using the 're-order hyperstack' function provided by Fiji. Labelled masks the entire area covered by the seed in each 2D slice were then created using Labkit, with the same label applied to all pixels contained within a single seed as it appeared across multiple slices. 3D masks of the entire shape (interior and exterior) of each seed were then created by stacking the slices with 2D label masks. During this annotation process the true number of seeds within each seed pod was recorded by manually counting the seeds within each 3D volume.

2.4 Seeds as star-convex polygons

To determine whether the shape of *B. napus* seeds could be appropriately described by star-convex polygons, the accuracy of reconstruction of ground truth labels for a small subset of 10 3D sub-volumes from the 'training' dataset was explored. Accuracy of reconstructed seeds was assessed based on the mean intersection-over-union (IoU) of ground-truth seed labels compared to 3D star-convex polyhedra shape representations of the seed, predicted using the method described by Weigert et al. (2020) in which for each pixel inside a seed the distance to the object boundary is calculated along a fixed set of rays that are approximately evenly distributed on an ellipsoid representative of the seeds within the dataset (see Weigert et al., 2020 eq. 1). The sets of rays used in seed reconstruction were calculated as Fibonacci rays, defined using the method described by Weigert et al. (2020), which were shown to be more accurate for reconstruction of 3D star-convex polyhedra than equidistant distributed rays and allowed for the potential anisotropy of seed to be taken into account. Reconstruction accuracy was investigated using a varying number of Fibonacci rays (8, 16, 32, 64, 96, and 128), as although Weigert et al. (2020) found a set of at least 64 rays was necessary to achieve accurate reconstruction of shape for cell nuclei, they suggested accurate reconstruction of less anisotropic or densely-packed objects may be possible with a smaller set of rays which would allow for less computational resources to be used in shape prediction.

2.5 Model training and validation

A StarDist-3D model with a U-Net backbone (Çicek et al., 2016) was trained to detect and segment individual *B. napus* seeds in 3D µCT sub-volumes from the labelled 'training' dataset using the pipeline described by Weigert et al. (2020). Model training was performed using a Google Colab runtime with 25.46 GB and a single GPU (Bisong, 2019). The StarDist-3D model was configured to use 96 Fibonacci rays in shape reconstruction, and to take into account the mean empirical anisotropy, of all labelled seeds in the dataset along each axis as calculated using the method described by Weigert et al., 2020 (X-axis = 1.103448275862069, Y-axis anisotropy = 1.032258064516129, Z-axis anisotropy = 1.0). The training patch size, referring to the size of the tiled portion of the 3D sub-volumes in the 'training' within view of the

neural network at any one time, was set to Z = 24, X= 96, and Y = 96 and training batch size set to 2. Training ran for 400 epochs with 100 steps per epoch and took 1.36 hours to complete (123ms/step).

Model validation was then performed by using the fine-tuned StarDist-3D algorithm to predict seed labels for all 3D µCT sub-volumes from the 'validation' dataset, which were then compared to the number and shape of seeds manually counted and labelled during annotation. Accuracy of seed detection and segmentation was then quantified for various levels of threshold τ, defined as the IoU between the predicted label and ground-truth label for each seed. The value of τ ranged between 0, where even a very slight overlap between predicted seeds and actual seeds counted as correctly predicted, and 1, where only predicted seed labels with pixel-perfect overlap with ground-truth labels counted as correctly predicted (Weigert et al., 2020).

Object detection accuracy was measured using the number of true positive results (TP), or number manually counted and labelled seeds that were correctly detected seeds, the number of false negative results (FN), or the number of manually counted and labelled seeds that were missed, the number of false positive results (FP), or number of objects other than seeds than were detected, recall, precision and F1-score. Recall related to the fraction of relevant objects that were successfully detected and was defined as:

$$Recall = \frac{TP}{TP + FN}$$

Precision related to the fraction of all detected objects that were relevant and was defined as:

$$Precision = \frac{TP}{TP + FP}$$

F1-score related to the harmonic mean of precision and recall, with the impact of precision and recall being given equal weight. F1-score was defined as:

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The accuracy of seed segmentation, or the accuracy of seed size and shape prediction, for the validation dataset was determined based on the mean matched score, defined as the mean IoU between the predicted and actual shape of true positive results, the mean true score, defined as the mean IoU between the predicted and actual shape of true positive results normalised by the total number of ground-truth labelled seeds, and panoptic quality, as defined in Eq.1 of Kirillov et al., 2019.

StarDist-3D models allow for specification of two values, the τ-threshold and the *nms*-threshold to optimize model output (Schmidt et al., 2018; Weigert et al., 2020). The τ-threshold refers to the minimum intersection-over-union between pairs of predicted and ground-truthed seeds required for detections to be classified as true positives, and can be set at 0.1 interval levels between 0.1 and 1 with 0.1 indicating a 10% overlap in the pixels within the predicted shape of a seed and the ground-truthed label and 1 respreseting a 100% overlap (Schmidt et al., 2018; Weigert et al., 2020). The *nms*-threshold, refers to the level of non-maximum suppression applied to the results of object detection and instance segmentation to prune the number of predicted star-convex polyhedra in ideally retain a single predicted shape for each true object, in this case each seed, within an image. The *nms*-threshold can be set at 0.1 interval levels between 0 and 1 with higher levels indicating more aggressive pruning of predicted shapes which therefore leads to fewer detections in the final model output. Therefore a higher *nms*-threshold is valuable in cases where the number of false positives expected in unfiltered model predictions is high. Both the τ-threshold and the *nms*-threshold for the fine-tuned StarDist-3D algorithm were set to optimal values based on the 'validation' dataset using the 'optimize_thresholds' function of StarDist (Schmidt et al., 2018).

2.6 Model testing and outputs

Testing of the fine-tuned StarDist-3D algorithm was carried out using the 'test' dataset, which was kept separate from model training and validation. Model testing was also carried out using the same Google Colab instance as model training and validation. Prediction, including both detection and segmentation of seeds took on average 1 minute 24 seconds to complete for a single complete 3D µCT volume containing a whole *B. napus* pod. Accuracy of seed detection was quantified using the same metrics as model validation, with the predicted number and location of seeds compared to the true number of locations of seeds in each image.

The output of prediction for the fine-tuned StarDist-3D algorithm were 3D numpy array volume containing labels depicting the predicted shape of seeds for each 3D µCT volume of a whole *B. napus* pod. The number and location of seeds including both bounding box and centroid coordinates on the Z, Y and X axis of the 3D volume could

TABLE 1 Example of data extracted on *Brassica napus* seed number per pod and seed location derived from fine-tuned StarDist-3D algorithm predictions.

Seed ID Number	Z-axis minimum	Y-axis minimum	X-axis minimum	Z-axis maximum	Y-axis maximum	X-axis maximum	Z-axis centroid	Y-axis centroid	X-axis centroid
1	354	258	221	387	293	257	370.3969	274.8411	238.3993
2	248	282	213	281	318	248	264.5435	299.7615	229.9509
3	504	264	214	538	299	250	520.8316	280.5683	231.2285
4	293	258	219	326	293	255	308.6688	274.7142	236.1587
5	185	271	206	215	304	239	199.2717	286.0577	221.4716

then be retrieved using the 'regionprops' and 'regionprops_table' functions of 'scikit-image' (van der Walt et al., 2014), an open-source python image analysis package (Table 1). Measurements of the 3D shape of seeds could also be extracted from the fine-tuned StarDist-3D model predictions using the 'regionprops_3D' and 'props_to_Dataframe' functions of 'porespy' (Gostick et al., 2019), an open-source python toolset for extracting data from 3D images of porous materials (Table 2). Seed size and shape metrics extracted using porespy functions included:

- Volume – the predicted volume of a detected seed in number of voxels
- Bounding box volume – the volume of the rectangular 3D bounding box containing a detected seed in number of voxels
- Sphericity – the ratio of the area of a sphere with the same volume as a detected seed to the predicted surface area of the same detected seed
- Surface area – the predicted surface area of a detected seed calculated using a reconstructed mesh of the surface contour of the seed
- Convex volume – number of pixels in the predicted convex hull image of a detected seed
- Equivalent diameter – the diameter of a circle with the same area as a detected seed
- Extent – ratio of pixels within the predicted shape of a detected seed to the total pixels within the 3D rectangular bounding box containing the seed
- Major axis length - the width of the thickest part of the seed, measured as a straight line through an ellipse that has the same normalized second central moments as the detected seed
- Minor axis length - the width of the thinnest part of the seed, measured as a straight line through an ellipse that has the same normalized second central moments as the detected seed
- Solidity – ratio of number of pixels within the predicted shape of a detected seed to number of pixels within the convex hull images of the same detected seed

Segmented images of individual detected seeds can also be exported. 3D volume images of individual seeds can be converted into numpy arrays and saved for further investigation using the open-source 'numpy' python package (Harris et al., 2020). As shown in Figure 1, individual 2D trans-axial slices showing a cross-section of detected seeds on both the XY and XZ axis can also be viewed and exported using the 'intensity_image' function of scikit-image (van der Walt et al., 2014).

2.7 Automated seed sorting by pod valve

In order to predict the valve in which the detected seeds were situated within the pod, coordinates of seeds detected with the fine-tuned StarDist-3D algorithm were converted to.csv format using the 'pandas' python package (McKinney, 2010) in order to allow loading into RStudio (RStudio Team, 2020). A locally weighted scatterplot smoothing (lowess) regression line was then fit to the XZ axis centroids of detected seeds using the 'lowess' function of the 'gplots' package in R (Cleveland 1979; Cleveland 1981; Warnes et al., 2005; R Core Team, 2018). The lowess regression line was then used to predict the division between the two valves of the seed pod, serving as a simplified reconstruction of the pod pseudoseptum, which is the membrane that separates both valves (Figure 2). The XZ centroid was used as all pods were arranged the same way during image collection so that the XZ plane displayed a cross-section of the pod with seeds sitting in one of two valves separated by the pod pseudoseptum, with the pod beak on the left and the pod pedicel on the right (Figure 3). In cases where pods contained less than or equal to 5 seeds, the smoother span (f), or proportion of points influencing the smooth at each value for the lowess regression line was set to $f = 1$. For seed pods containing greater than 5 seeds the default value of f provide by the 'lowess' function was used. The vertical distance between the XZ centroid of detected seeds and lowess regression line was then calculated and seeds found to be above the lowess regression line were determined to belong to 'valve 1' while seeds below the line were determined to belong to 'valve 2'. The sequence number for detected seeds in each valve from pod beak to pedicel, and the distance between sequential seeds in each valve could then be calculated and added to the.csv data of seed coordinates for each seed pod. The base R function 'for' was used to create a looping script to automate the prediction of valve and calculation of valve related metrics for all seeds in all seedpods and on average it took 70 milliseconds to complete valve

TABLE 2 Example of data extracted on *Brassica napus* seed size and shape derived from fine-tuned StarDist-3D algorithm predictions.

Seed ID Number	Volume	Bounding box volume	Sphericity	Surface area	Convex volume	Equivalent diameter	Extent	Major axis length	Minor axis length	Solidity
1	23237	41580	0.886865	4440.313	24538	35.40466	0.55885	45.38723	44.23421	0.94698
2	23379	41580	0.89321	4426.713	24627	35.47664	0.562266	45.43497	44.46307	0.949324
3	23344	42840	0.870115	4539.668	24680	35.45892	0.544911	45.77541	44.01394	0.945867
4	21921	41580	0.82114	4612.895	23238	34.72327	0.527201	44.78899	43.01572	0.943326
5	19451	32670	0.95953	3645.191	20476	33.36679	0.595378	43.07271	41.68581	0.949941

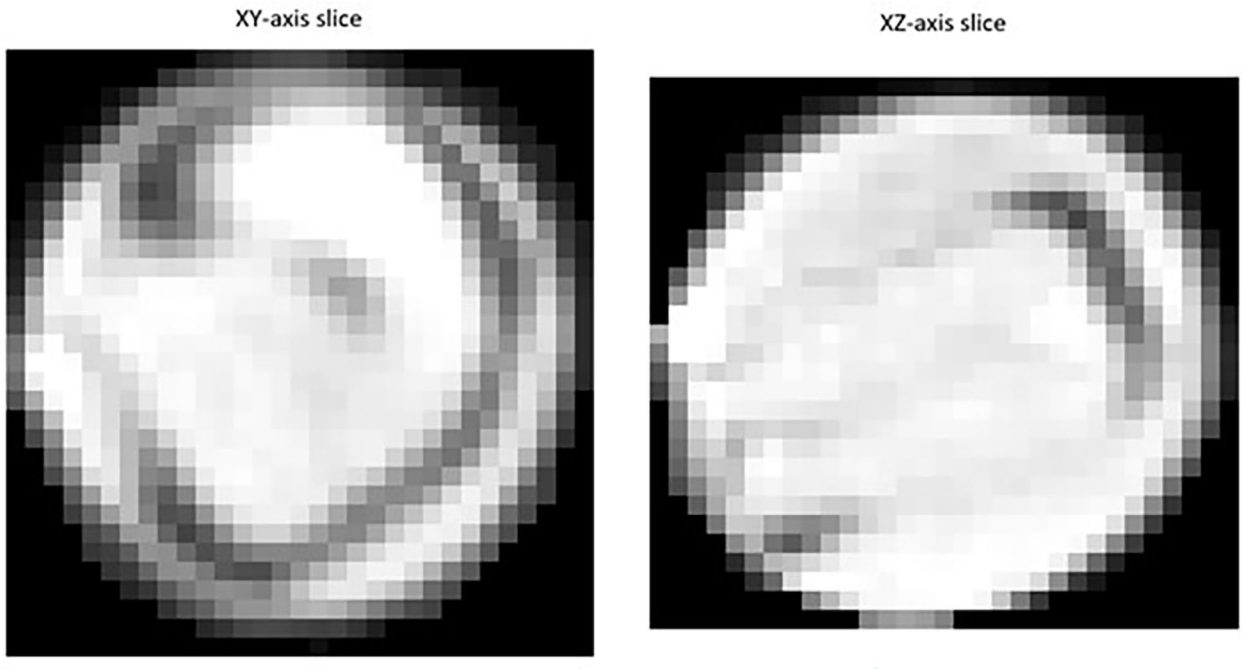


FIGURE 1
Example of a 2D slice images extracted for individual *Brassica napus* seeds detected and segmented with a fine-tuned StarDist-3D algorithm from 3D micro-computed tomography.

prediction and valve-related metric extraction for an entire pod using a single CPU.

3 Results

3.1 Accuracy of reconstruction of seed labels as star-convex polygons

As shown in Figure 5, sufficiently accurate reconstruction (greater than 0.8 mean IoU) of labelled seeds was achieved with as few as 32 rays with or without taking anisotropy into account, and the highest reconstruction accuracy (greater than 0.9 mean IoU) was achieved when 64 rays or more were used. It was therefore determined that it was appropriate to describe the shape of seeds as star-convex polyhedral and to proceed with training a StarDist-3D for detection and segmentation of seeds. It was also decided that reconstruction with anisotropy should be used that would more easily allow application of the workflow described in this paper to images of seeds or other star convex plant structures that may be more irregular in shape. Example reconstruction of seed shape with anisotropy taken into account is demonstrated in Figure 6. See Supplementary Figure 1 for seed reconstruction without anisotropy.

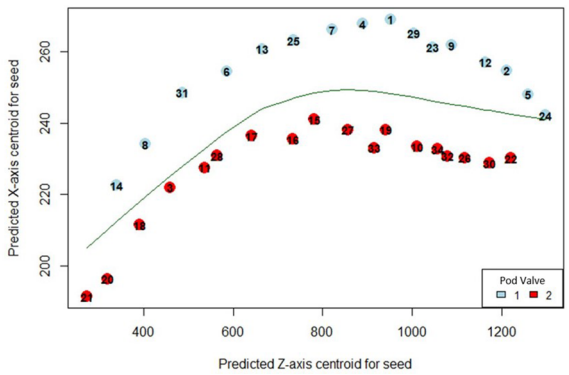


FIGURE 2
Predicted position of automatically detected *Brassica napus* seeds in pod valves using automated lowess regression. Points on the graph indicate the XZ centroid of detected seeds with a unique seed identification number and are coloured based on whether they were predicted to be positioned in valve 1 (blue) or valve 2 (red).

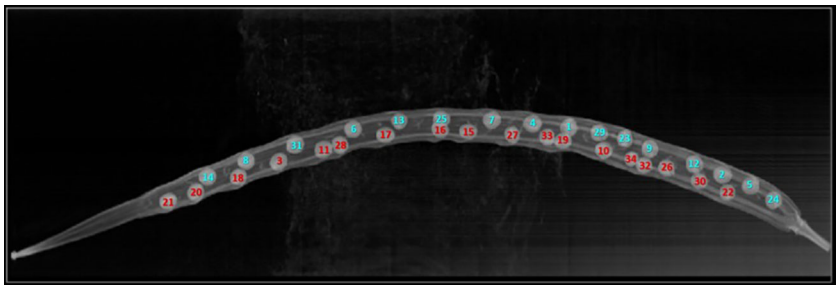


FIGURE 3
Original XZ slice image of a *Brassica napus* pod (same pod for which valve predicted was performed as shown in Figure 4). Seeds are marked with a unique identification number matching Figure 4 and are coloured by whether they were confirmed to be positioned in valve 1 (blue) or valve 2 (red) through manual analysis.

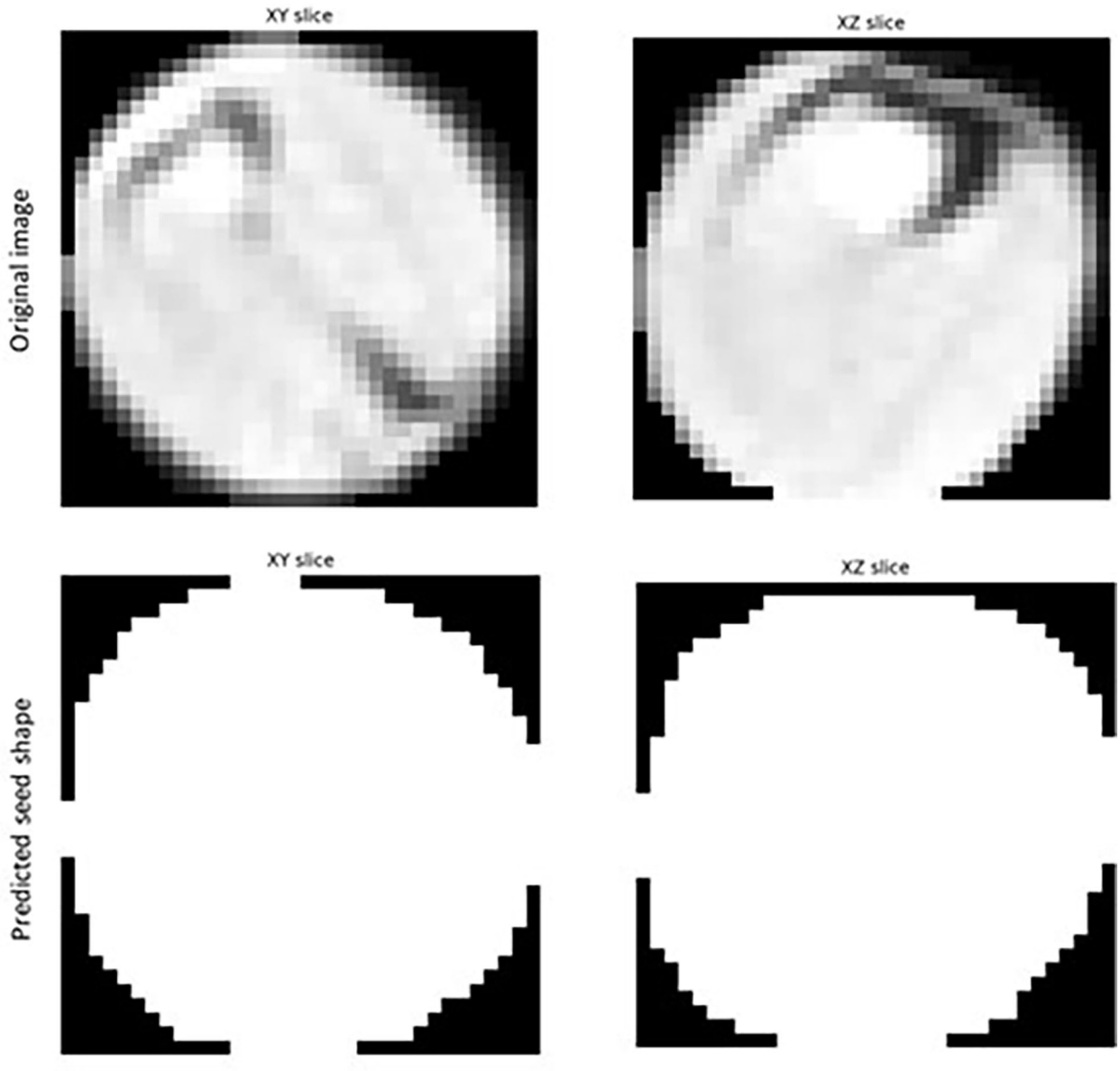


FIGURE 4
Example of segmentation results for individual *Brassica napus* seeds detected in 3D micro-computed tomography scans of seed pods from the 'test' dataset.

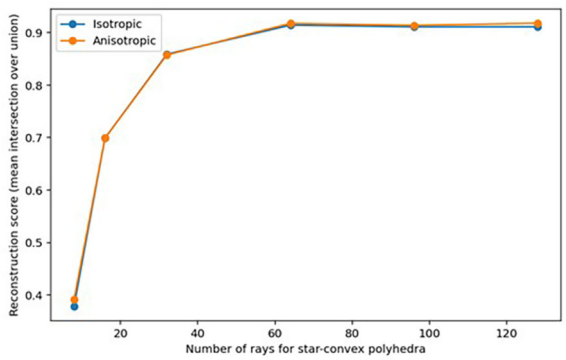


FIGURE 5
Reconstruction accuracy (mean IoU) of ground-truth labelled *Brassica napus* seeds when using different unit Fibonacci rays.

3.2 Seed detection and segmentation model validation

The trained StarDist-3D model was tested with different certainty thresholds (τ) to predict the number and shape of seeds within a pod. For τ = 0.1 to 0.8, 39 of the 42 seeds contained within the validation data were detected using the fine-tuned StarDist-3D algorithm and there were no false positive results (Table 3). Therefore, the recall rate (the actual number of seeds in the image that were successfully detected) was 92.9%, the precision rate (the number of detected objects in the image that were seeds) was 100%, and F1-score was 96.3% for the validation dataset across this range of τ (Table 3).

The mean matched score when τ = 0.1 to 0.8 was 0.900, indicating a 90.0% overlap in pixels predicted to be a part of detected seeds with pixels known to be a part of ground-truth labelled seeds (Table 3). The mean true score (0.836) and panoptic quality (0.867) for this range of τ also suggested a high degree of overlap between the predicted and actual shape of detected seeds (Table 3). Figure 7 displays example segmentation results for individual seeds detected with the fine-tuned StarDist-3D algorithm.

When τ was increased from 0.8 to 0.9 a slight increase in mean matched score from 0.900 to 0.912 occurred, but a large decrease in accuracy of both detection and segmentation as indicated by all other metrics was observed (Table 3). Setting thresholds of τ = 0.7 and *nms* = 0.4 resulted in the highest precision, recall and F1-score accuracy for seed detection and were therefore identified as optimal and incorporated into the fine-tuned StarDist-3D model used to perform prediction on novel pod data.

3.3 Seed detection and segmentation model testing

The true number of seeds contained within 'test' dataset was 209, while the total number of seeds predicted to be present within the 'test' images using the fine-tuned StarDist-3D algorithm was 208 (Table 4). One predicted seed was determined to be a pod pedicel incorrectly labelled as a seed and was recorded as a false positive result, while two seeds observed in the 'test' dataset were missed and recorded as false negative results. Therefore, the overall precision rate of the fine-tuned StarDist-3D algorithm when applied to the 'test' dataset was 99.52%, the overall recall rate was 99.04%, and the overall F-score was 99.28% (Table 4). Within individual pods the precision rate ranged from 95-100%, the recall rate ranged from 90.48-100%, and the F-score ranged from 95-100% (Table 4).

Example segmentation results for individual seeds detected in pod images from the 'test' dataset are displayed in Figure 4. A large degree of variance in the appearance of the interior of detected and segmented seeds, in particular the amount of empty space within

TABLE 3 Accuracy metrics for automated detection and segmentation of *Brassica napus* seeds in 3D micro-computed tomography scans from the 'validation' dataset across several intersection-over-union thresholds τ.

Threshold τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Detection Metrics									
True Number of Seeds	*42*	*42*	*42*	*42*	*42*	*42*	*42*	*42*	*42*
Number of True Positives	*39*	*39*	*39*	*39*	*39*	*39*	*39*	*39*	*25*
Number of False Negatives	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*17*
Number of False Positives	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*14*
Precision	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*1.000*	*0.641*
Recall	*0.929*	*0.929*	*0.929*	*0.929*	*0.929*	*0.929*	*0.929*	*0.929*	*0.595*
F1-score	*0.963*	*0.963*	*0.963*	*0.963*	*0.963*	*0.963*	*0.963*	*0.963*	*0.617*
Segmentation Metrics									
Mean true score	*0.836*	*0.836*	*0.836*	*0.836*	*0.836*	*0.836*	*0.836*	*0.836*	*0.543*
Mean matched score	*0.900*	*0.900*	*0.900*	*0.900*	*0.900*	*0.900*	*0.900*	*0.900*	*0.912*
Panoptic quality	*0.867*	*0.867*	*0.867*	*0.867*	*0.867*	*0.867*	*0.867*	*0.867*	*0.563*

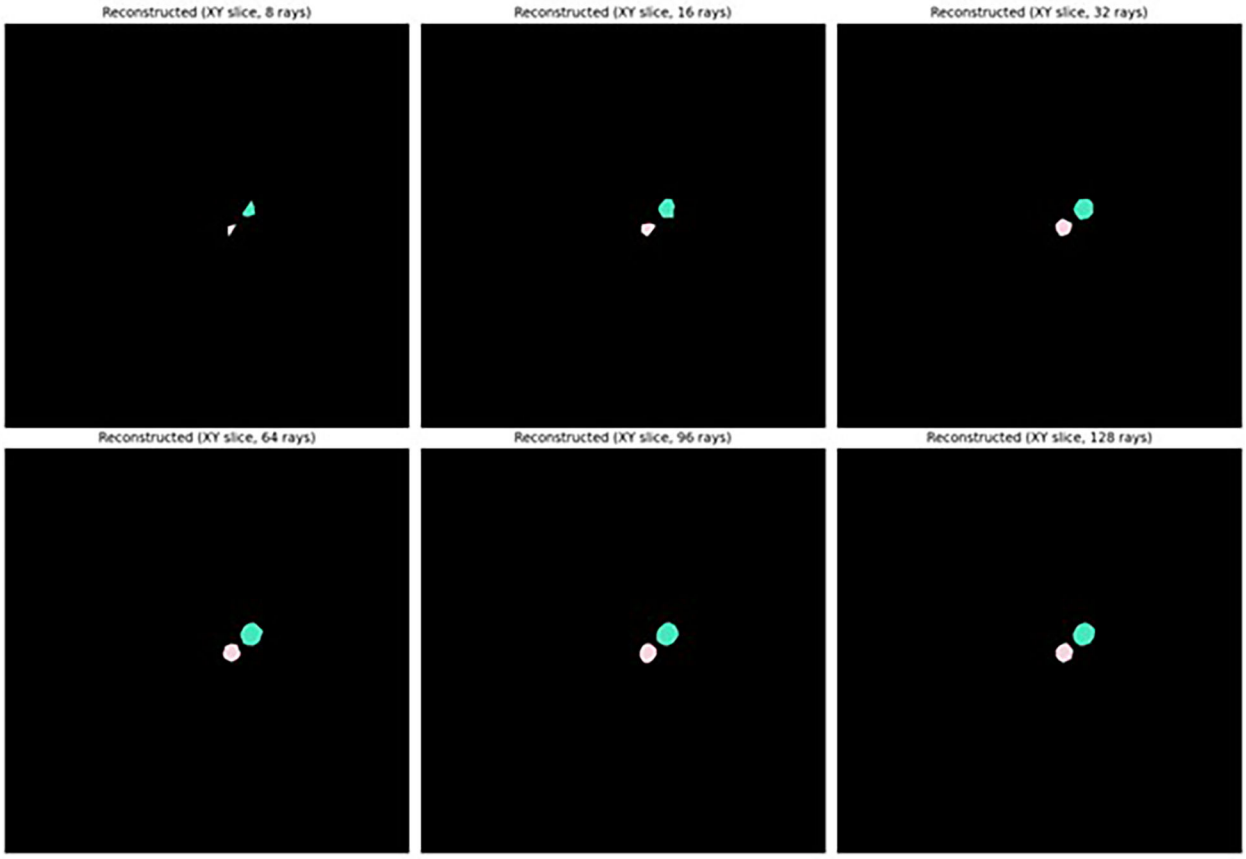


FIGURE 6
Reconstructed shape of *Brassica napus* seeds using different numbers of Fibonacci rays with anisotropy of seeds taken into account.

the seed (depicted by pixels closer to black in colour) was observed, even within the same pod (Figure 8).

3.4 Automated seed sorting by valve

198 out of the 209 seeds contained in the 'test' dataset were sorted into the correct valve using the automated seed sorting method, resulting in an overall accuracy of 94.74% (Table 5). The percentage of correctly sorted seeds within a single pod ranged from 50-100%, with seed sorting accuracy below 85.71 only occurring for pods that contained 6 or fewer total seeds (Table 5).

4 Discussion

The overall accuracy of the StarDist-3D model fine-tuned on 3D µCT images of *B. napus* pods was higher than reported for the developmental use case of detection and segmentation of individual cell nuclei (Weigert et al., 2020). This may be due to the very high resolution of 3D µCT images, and smaller number of target objects within them, which meant that individual seeds, as opposed to cell nuclei, were less likely to be obscured, and there were few cases of closely clustered target objects. The high precision of the fine-tuned StarDist-3D model may also be due to the fact that no other pod structures within the images closely resemble seeds in shape. This demonstrates the suitability of the 3D µCT *B. napus* pod image dataset to automated data extraction. A small number of false negative errors occurred when seeds were much smaller and unevenly shaped compared to the majority of seeds, possibly due to post-fertilization seed abortion. Dissection of the pods to determine the cause of this size and shape disparity in these missed seeds, and further fine-tuning of the pre-trained StarDist-3D model based on a curated dataset of seeds of more diverse size and shape is likely to improve the recall rate. However, the low error rate already achieved with a relatively small amount of annotated training data suggests the high resolution of the 3D µCT *B. napus* pod images make them a valuable resource highly compatible with state-of-art computer vision approaches.

The fine-tuned StarDist-3D approach utilized in this paper allowed for accurate data on the number, spatial arrangement, size and shape of seeds to be extracted from a 3D µCT image of a whole *B. napus* pod in under 1 minute 30 seconds, as opposed to manual image analysis methods which took approximately 14 minutes to obtain only a small subset of the measurements. Manual methods are more labour intensive as pods have to be collected and placed flat with a contrasting background to obtain high-quality images for semi-automatically measuring valve and beak length by using SmartRoot tool in Fiji (Schindelin et al., 2012). Then, the pods have to be manually opened to obtain SNPP data. The manual

TABLE 4 Accuracy metrics for automated detection of *Brassica napus* seeds in 3D micro-computed tomography scans of pods from the 'test' dataset.

Seed Pod ID	True Number of Seeds	Predicted Number of Seeds	Number of True Positives	Number of False Positives	Number of False Negatives	Recall	Precision	F1-score
C0007186	5	5	5	0	0	100	100	100
C0007197	19	20	19	1	0	100	95	97.44
C0007198	4	4	4	0	0	100	100	100
C0007205	2	2	2	0	0	100	100	100
C0007224	34	34	34	0	0	100	100	100
C0007226	6	6	6	0	0	100	100	100
C0007239	8	8	8	0	0	100	100	100
C0007256	30	30	30	0	0	100	100	100
C0007262	2	2	2	0	0	100	100	100
C0007269	1	1	1	0	0	100	100	100
C0007274	34	34	34	0	0	100	100	100
C0007299	3	3	3	0	0	100	100	100
C0007311	15	15	15	0	0	100	100	100
C0007420	7	7	7	0	0	100	100	100
C0007440	21	19	19	0	2	90.48	100	95
C0007456	4	4	4	0	0	100	100	100
C0007853	6	6	6	0	0	100	100	100
C0007864	4	4	4	0	0	100	100	100
C0007865	4	4	4	0	0	100	100	100
Total	**209**	**208**	**207**	**1**	**2**	**99.04**	**99.52**	**99.28**

process scan can result in loss of seeds and inaccurate counting. Opening the pod invariably leads to movement of the seed within the pod, therefore information on the valve and spacing is lost. Hence, the automated pipeline that we describe here is more efficient and less time consuming than currently used methods. In addition, the automatic seed sorting by valve step also allowed for data to be collected on the spatial arrangement of seeds in relation to both each other and pod valves that cannot be examined through conventional dissection methods of examining pods. For pods with a very low number of seeds (six or less) valve misclassification errors were more common, demonstrating that it was difficult to reliably predict the position and shape of pseudoseptum from a small number of data points on seed location. Future development of an additional edge detection algorithm to directly detect the pseudoseptum, rather than relying solely on seed position, could improve the accuracy of valve sorting for pods with very few seeds. However, this study demonstrates the significantly less computationally intensive graphical method of predicting pseudoseptum shape and seed valve position using lowess regression is suitable for analysis of images of pods with 7 seeds or more.

The reduction in the bottleneck for analysing the 3D µCT image dataset provided by the StarDist-3D approach could enable detailed data on the number, size, shape and spatial arrangement of seeds to be integrated into models of plant development. It could also potentially be applied to 3D µCT imagery of seeds from other species or other plant structures with relatively little retraining effort provided the target objects can appropriately be described as star-convex polygons. The gradient of seed growth within a pod, the difference of seed growth and abortion within pods in different positions in the main inflorescence, comparison of pods between the main and the secondary inflorescences and the effect of different environmental perturbations, such as heat stress, could be studied. Hence, this method of analysis that does not require opening pods will help to better understand SNPP and seed abortion and their relation to plant seed yield in several crops. A multi-class version of the StarDist-3D model could be trained to predict shape and position of beaks, pedicels and post-fertilisation aborted seeds as these structures can also be accurately reconstructed as star-convex polygons. This will allow the automatically extraction of further metrics such as beak length and overall pod length, as well as reduce the potential of false negative errors caused by aborted seeds that tend to be unusually small and irregularly shaped compared to mature seeds. This extension would rely upon annotating a larger number of full 3D µCT as each only contains a single beak and pedicel, however, since accurate results were obtained with the number of seeds labelled for training data in this study researchers

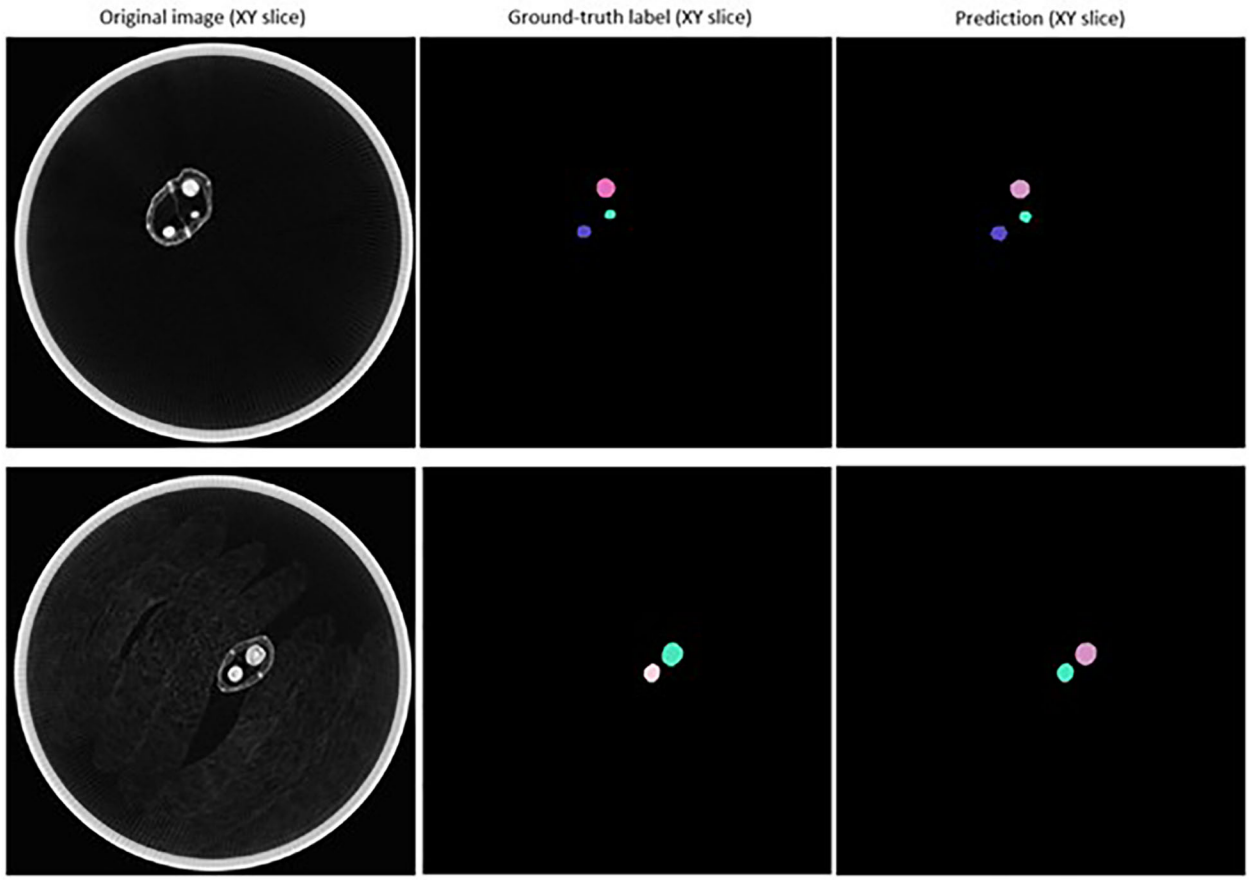


FIGURE 7
Example of segmentation results for *Brassica napus* seeds detected in 3D micro-computed tomography scans of seed pods from the 'validation' dataset (t = 0.7).

may only need to label beaks and pedicels in further imagery in order to prepare an adequate multi-class training dataset.

Scale has been shown to have negligible impact on accuracy of StarDist-3D object detection and segmentation, therefore the fine-tuned model described could reliably be applied to images of B. napus seed pods that vary significantly in size, both in terms of the pods themselves and size of individual seeds. The findings of this study also demonstrated that the Stardist-3D method could be applied to imagery of seed from other plant species, as well as other plant structures such as peas, nuts or grains that are rounded or ovate in shape and therefore can be accurately reconstructed as star-convex polyhedra. This potentially includes more unusually shaped, less spherical seeds and nuts from species such as *Arabidopsis thaliana*, *Camelina sativa* and *Arachis hypogaea* (peanuts) as the anisotropy, or non-spherical irregularities of these seeds, could be taken into account with the Stardist-3D method.

The fine-tuned StarDist-3D model can also likely be reliably applied to datasets with lower contrast than the Scanco μCT images used in this study, as the StarDist-3D methods have been demonstrated to yield more accurate results for low contrast, low signal-to-noise 3D volume data compared to other contemporary deep learning based instance segmentation methods (Schmidt et al., 2018; Weigert et al., 2020). It is recommended that data augmentation methods be applied to the study dataset in future to explore the effect of resolution on performance, as at a sufficiently low resolution the accuracy of shape prediction may be impacted due to the borders of seeds being blurred (Schmidt et al., 2018; Weigert et al., 2020).

The automated clustering of the segmented images of individual seeds that are output by the fine-tuned StarDist-3D with a rotationally invariant method can also be explored, as a high degree of variation was observed in the internal structure of the seeds. The segmented seed images would need to be represented as rotationally invariant images in order to explore clustering, as seeds are oriented at different directions within pods and rotationally invariant representation would negate the effect of these differing orientations so that other similarities and differences in the internal appearance of the seeds could be quantified (Zhao & Singer 2014). This is a very promising step once images with higher resolution are acquired. Clustering could reveal similarities in internal appearance between groups of seeds that could be linked to biological origins, which could be ground-truthed through manual examination and seed dissection after 3D μCT. Moreover, synchrony and different orientation of the seeds could be further explored. This knowledge is of high importance as breeders pursue good and synchrony of seed maturation. Therefore, links between

10.3389/fpls.2023.1120182

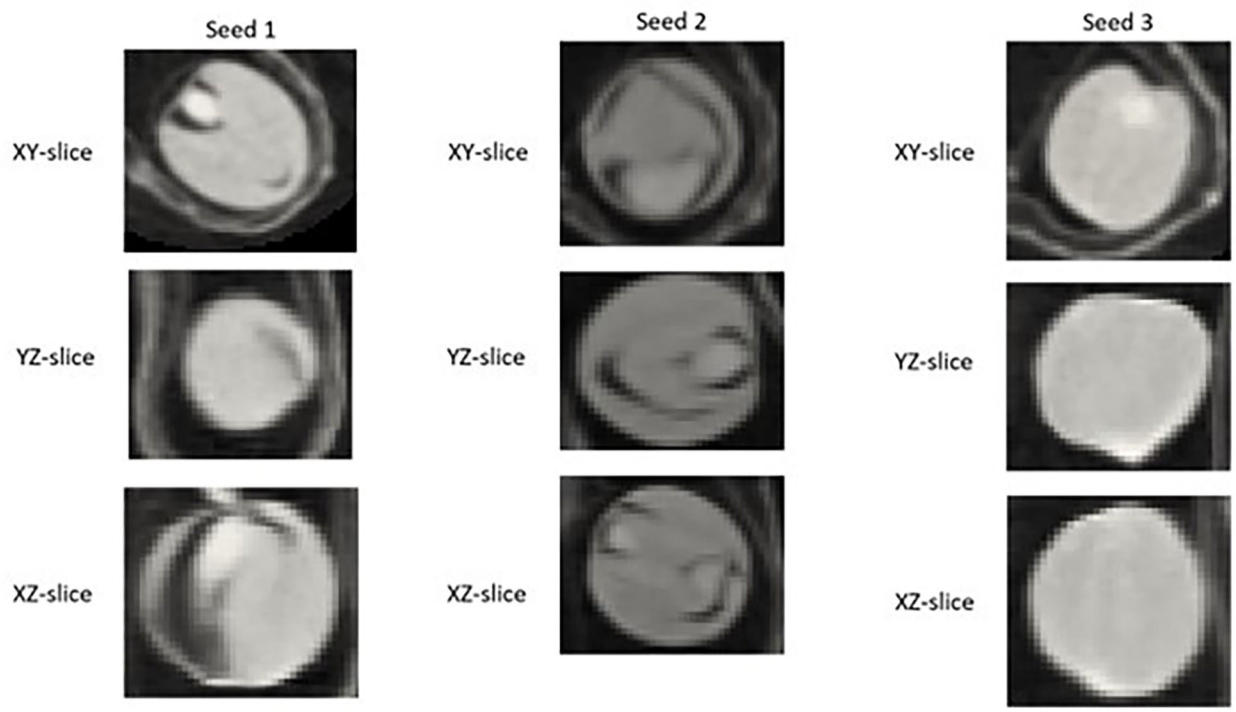


FIGURE 8
2D XY, YZ, and XZ slice images of individual *Brassica napus* seeds detected and segmented with a fine-tuned StarDist-3D algorithm from the same pod displaying variation in internal structure, particularly empty (black) space inside of seeds.

TABLE 5 Number and percentage of *Brassica napus* seeds automatically sorted into the correct pod valve.

Seed Pod ID	Valve 1		Valve 2		Number of Correctly Sorted Seeds	Percentage of Correctly Sorted Seeds (%)
	True Number of Seeds	Predicted Number of Seeds	True Number of Seeds	Predicted Number of Seeds		
C0007186	4	5	1	0	4	80
C0007197	9	10	10	10	19	100
C0007198	3	3	1	1	4	100
C0007205	1	2	1	0	1	50
C0007224	16	16	18	18	34	100
C0007226	3	3	3	3	6	100
C0007239	2	3	6	5	7	87.5
C0007256	17	17	13	13	30	100
C0007262	2	2	0	0	2	100
C0007269	0	0	1	1	1	100
C0007274	18	19	16	15	33	97.06
C0007299	0	0	3	3	3	100
C0007311	8	8	7	7	15	100
C0007420	5	4	2	3	6	85.71
C0007440	11	12	10	7	18	85.71

(Continued)

TABLE 5 Continued

Seed Pod ID	Valve 1		Valve 2		Number of Correctly Sorted Seeds	Percentage of Correctly Sorted Seeds (%)
	True Number of Seeds	Predicted Number of Seeds	True Number of Seeds	Predicted Number of Seeds		
C0007456	3	2	1	2	3	75
C0007853	4	3	2	3	5	83.33
C0007864	1	1	3	3	4	100
C0007865	1	2	3	2	3	75
Total	**108**	**112**	**101**	**96**	**198**	**94.74**

these traits and the growing conditions and genotypes of the plants that the seeds were collected from can be explored in order to better understand factors affecting seed maturation and plant yield.

5 Conclusion

High-throughput plant image datasets have the potential to greatly improve our understanding of the factors affecting dynamic responses in plant development and crop yield, however a lack of reliable and efficient methods for extracting phenotype data from these datasets remains a major bottleneck. This paper demonstrates that an existing, state-of-the-art object detection and segmentation method, StarDist-3D, can be applied with little modification to automatically obtain seed number, size, shape and spatial seed arrangement from 3D µCT images of *B. napus* pods in a time-saving and non-destructive manner with a high degree of accuracy. This method could enable the study of seed development within a specific time-point or during different phases of pod growth, obtaining very specific and detailed information that otherwise, would not be possible to accurately capture. Acquiring information regarding the internal structure of opaque pods can be incorporated into high-throughput plant phenotyping platforms and enables the opportunity of understanding and linking pod, seed development and disposition within different germplasm and plant developmental responses to biotic and abiotic stresses. The findings of this paper also demonstrate how current gold standard computer vision methods can be generalised to accurately analyse imagery collected using a variety of sensors, at different scales and from a wide range of scientific domains.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

EC annotated training data, led development of the automated plant phenotyping pipeline, analyzed results of the pipeline, and wrote the first draft. LS and SK collected the plant image datasets. LS, SK, and SA contributed to scoping the developing the automated plant phenotyping pipeline and reviewing and editing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/W006022/1, particularly the "Environment and Sustainability" theme within that grant & The Alan Turing Institute. LS and SK were supported by the UK Biotechnology and Biological Sciences Research Council (grant BBS/E/C/000I0420 and BB/P003095/1).

Acknowledgments

The authors want to thank the plant phenomics facility in Aberystwyth University (Wales, UK) for the obtention of the µCT scanner images of the mature *Brassica napus* pods.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1120182/full#supplementary-material

References

Alalwan, N., Abozeid, A., ElHabshy, A. A., and Alzahrani, A. (2021). Efficient 3D deep learning model for medical image semantic segmentation. *Alexandria Engineering Journal (AEJ)*, 60 (1), 1231–1239. doi: 10.1016/j.aej.2020.10.046.

Alseekh, S., Kostova, D., Bulut, M., and Fernie, A. R. (2021). Genome-wide association studies: Assessing trait characteristics in model and crop plants. *Cell. Mol. Life Sci.* 78, 5743–5754. doi: 10.1007/s00018-021-03868-w

Arzt, M., Deschamps, J., Schmied, C., Pietzsch, T., Schmidt, D., Tomancak, P., et al. (2022). LABKIT: Labeling and segmentation toolkit for big image data. *Front. Comput. Sci.* 4. doi: 10.3389/fcomp.2022.777728

Beucher, S., and Meyer, F. (1993). The morphological approach to segmentation: The watershed transformation. In E. Dougherty (Ed.), *Mathematical morphology in image processing* (pp. 433–481). Marcel Dekker. doi: 10.1201/9781482277234-12

Bisong, E. (2019). "Google Colaboratory," in *Building machine learning and deep learning models on Google cloud platform* (Berkeley, CA: Apress). doi: 10.1007/978-1-4842-4470-8_7

Bours, R., Muthuraman, M., Bouwmeester, H., and van der Krol, A. (2012). OSCILLATOR: A system for analysis of diurnal leaf growth using infrared photography combined with wavelet transformation. *Plant Methods* 8, 29. doi: 10.1186/1746-4811-8-29

Boykov, Y., and Funka-Lea, G. (2006). Graph cuts and efficient ND image segmentation. *Int. J. Comput. Vis.* 70:2, 109–131. doi: 10.1007/s11263-006-7934-5

Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biol.* 12, 232. doi: 10.1186/gb-2011-12-10-232

Caicedo, J. C., Roth, J., Goodman, A., Becker, T., Karhohs, K. W., Broisin, M., et al. (2019). Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry A* 95:9, 952–965. doi: 10.1002/cyto.a.23863

Cheng, J., and Rajapakse, J. C. (2009). Segmentation of clustered nuclei with shape markers and marking function. *IEEE Trans. Biomed. Eng.* 56:3, 741–748. doi: 10.1109/TBME.2008.2008635

Çicek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). *3D U-net: learning dense volumetric segmentation from sparse annotation* (Athens, Greece: Medical Image Computing and Computer-Assisted Intervention (MICCAI), 424–432. doi: 10.1007/978-3-319-46723-8_49

Çiçek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). *3D U-net: Learning dense volumetric segmentation from sparse annotation* (Athens, Greece: MICCAI), 2016.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Am Stat Assoc* 74 (368), 829–836. doi: 10.1080/01621459.1979.10481038.

Cleveland, W. S. (1981). LOWESS: A program for smoothing scatter plots by robust locally weighted regression. *Am. Stat* 35, 54. doi: 10.2307/2683591

Costa, C., Schurr, U., Loreto, F., Menesatti, P., and Carpentier, S. (2019). Plant phenotyping research trends, a science mapping approach. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01933

Dhondt, S., Vanhaeren, H., Van Loo, D., Cnudde, V., and Inze, D. (2010). Plant structure visualization by high-resolution X-ray computed tomography. *Trends Plant Sci.* 15:8, 414–422. doi: 10.1016/j.tplants.2010.05.002

Domander, R., Felder, A. A., and Doube, M. (2021). BoneJ2 - refactoring established research software. *Wellcome Open Res.* 6, 37. doi: 10.12688/wellcomeopenres.16619.2

Easlon, H. M., and Bloom, A. J. (2014). Easy leaf area: Automated digital image analysis for rapid and accurate measurement of leaf area. *Appl. Plant Sci.* 2:7, 1400033. doi: 10.3732/apps.1400033

Gargiulo, L., Leonarduzzi, C., and Mele, G. (2020). Micro-CT imaging of tomato seeds: Predictive potential of 3D morphometry on germination. *Biosyst. Eng.* 200,200. doi: 10.1016/j.biosystemseng.2020.09.003

Gostick, J., Khan, Z. A., Tranter, T. G., Kok, M. D. R., Agnaou, M., Sadeghi, M. A., et al. (2019). PoreSpy: A Python toolkit for quantitative analysis of porous media images. *J. Open Source Softw.* 4(37), 1296. doi: 10.21105/joss.01296

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi: 10.1038/s41586-020-2649-2

He, K., Gkioxari, G., Dollár, P., Girshick, R., and Mask, R. C. N. N. (2017) (Accessed IEEE International Conference on Computer Vision (ICCV)).

Hu, W., Zhang, C., Jiang, Y., Huang, C., Liu, Q., Xiong, L., et al. (2020). Nondestructive 3D image analysis pipeline to extract rice grain. *Plant Phenomics* 2020, 3414926. doi: 10.34133/2020/3414926

Hughes, A., Askew, K., Scotson, C. P., Williams, K., Sauze, C., Corke, F., et al. (2017). Non-destructive, high-content analysis of wheat grain traits using X-ray micro computed tomography. *Plant Methods* 13, 76. doi: 10.1186/s13007-017-0229-8

Kar, J., Cohen, M. V., McQuiston, S. P., and Malozzi, C. M. (2021). A deep-learning semantic segmentation approach to fully automated MRI-based left-ventricular deformation analysis in cardiotoxicity. *Magn Reson Imaging* 78, 127–139. doi: 10.1016/j.mri.2021.01.005

Kirillov, A., He, K., Girshick, R., Rother, C., and Dollar, P. (2019) *Panoptic segmentation*. Available at: https://arxiv.org/abs/1801.00868.

Kunishima, N., Takeda, Y., Hirose, R., Kalasova, D., Salplachta, J., and Omote, K. (2020). Visualization of internal 3D structure of small live seed on germination by laboratory-based X-ray microscopy with phase contrast computed tomography. *Plant Methods* 16, 7. doi: 10.1186/s13007-020-0557-y

Lin, W. J., Chen, J. W., Jhuang, J. P., Tsai, M. S., , HUng C.L., Li, K. M., et al. (2021). Integrating object detection and image segmentation for detecting the tool wear area on stitched image. *Sci. Rep.* 11, 19938. doi: 10.1038/s41598-021-97610-y

Liu, W., Liu, C., Jin, J., Li, D., Fu, Y., and Yuan., X. (2020). High-throughput phenotyping of morphological seed and fruit characteristics using X-ray computed tomography. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.601475

Lotufo, R. A., Falcao, A. X., and Zampirolli, F. A. (2002). "IFT-watershed from gray-scale marker," in *Proceedings. XV Brazilian Symposium on Computer Graphics and Image Processing*, Fortaleza, Brazil, 2002, pp. 146–152. doi: 10.1109/SIBGRA.2002.1167137.

Majanga, V., and Viriri, S. (2021). Dental images' segmentation using threshold connected component analysis. *Comput. Intell. Neurosci.* 2021, 2921508. doi: 10.1155/2021/2921508

Mathers, A. W., Hepworth, C., Baillie, A. L., Sloan, J., Jones, H., Lundgren, M., et al. (2018). Investigating the microstructure of plant leaves in 3D with lab-based X-ray computed tomography. *Plant Methods* 14, 99. doi: 10.1186/s13007-018-0367-7

McKinney, W. (2010)Data structures for statistical computing in python (Accessed Proceedings of the 9th Python in Science Conference).

Meijering, E. (2012). Cell segmentations: 50 years down the road. *IEEE Signal Process. Magazine* 29:5, 140–145. doi: 10.1109/MSP.2012.2204190

Narisetti, N., Neumann, K., Röder, M. S., and Gladilin, E. (2020). Automated spike detection in diverse European wheat plants using textural features and the frangi filter in 2D greenhouse images. *Front. Plant Sci.* 23:11. doi: 10.3389/fpls.2020.00666

Pajor, R., Fleming, C., Osborne, C. P., Rolfe, S. A., Sturrock, C. J., and Mooney, S. J. (2013). Seeing space: Visualization and quantification of plant leaf structure using X-ray micro-computed tomography: View point. *J. Exp. Bot.* 64:2, 385–390. doi: 10.1093/jxb/ers392

Piovesan, A., Vancauwenberghe, V., Van De Looverbosch, T., Verboven, P., and Nicolai, B. (2021). X-Ray computed tomography for 3D plant imaging. *Trends Plant Sci.* 11, 1171–1185. doi: 10.1016/j.tplants.2021.07.010

R Core Team (2018). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/.

Rousseau, D., Widiez, T., Di Tommaso, S., Rositi, H., Adrien, J., Maire, E., et al. (2015). Fast virtual histology using X-ray in-line phase tomography: application to the 3D anatomy of maize developing seeds. *Plant Methods* 11, 55. doi: 10.1186/s13007-015-0098-y

RStudio Team (2020). *RStudio: Integrated development for r* (Boston, MA: RStudio, PBC). Available at: http://www.rstudio.com/.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9:7, 676–682. doi: 10.1038/nmeth.2019

Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018)Cell detection with star-convex polygons (Accessed International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)).

Schneider, J. V., Rabenstein, R., Wesenberg, J., Karsten, W., Zizka, G., and Habersetzer, J. (2018). Improved non-destructive 2D and 3D X-ray imaging of leaf venation. *Plant Methods* 14, 7. doi: 10.1186/s13007-018-0274-y

Seethepalli, A., Dhakal, K., Griffiths, M., Guo, H., Freschet, G. T., and York, L. M. (2021). RhizoVision explorer: Open-source software for root image analysis and measurement standardization. *Ann. Bot.* 13:6. doi: 10.1093/aobpla/plab056

Siles, L., Hassall, K. L., Sanchhis-Gritsch, C., Eastmond, P. J., and Kurup, S. (2021). Uncovering trait associations resulting in maximal seed yield in winter and spring oilseed rape. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.697576

Sodjinou, S., Mohammadi, V., Mahama, A. T.S., and Gouton, P. (2022). A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images. *Inf. Process. Agric*, 9(3), 355–364. doi: 10.1016/j.inpa.2021.08.003

Tracy, S. R., Gómez, J. F., Sturrock, C. J., Wilson, Z. A., and Ferguson, A. C. (2017). Non-destructive determination of floral staging in cereals using X-ray micro computed tomography (μCT). *Plant Methods* 13, 9. doi: 10.1186/s13007-017-0162-x

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., et al. (2014). Scikit-image: Image processing in Python. *PeerJ* 2, e453. doi: 10.7717/peerj.453

Van Valen, D. A., Kudo, T., Lane, K. M., Macklin, D. N., Quach, N. T., DeFelice, M. M., et al. (2016). Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PloS Comput. Biol.* 12:11, e1005177. doi: 10.1371/journal.pcbi.1005177

Wang, Y., Ahsan, U., Li, H., and Hagen, M. (2022). A comprehensive review of modern object segmentation approaches. *Trends Comput* 13 (2–3), 111–283. doi: 10.1561/0600000097

Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., et al. (2005) *Gplots: Various r programming tools for plotting data*. Available at: https://cran.r-project.org/package=gplots.

Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (2020)Star-convex polyhedra for 3D object detection and segmentation in microscopy (Accessed The IEEE Winter Conference on Applications of Computer Vision (WACV)).

Xie, W., Noble, J. A., and Zisserman, A. (2018). Microscopy cell counting and detection with fully convolutional regression networks. *Comput. Methods Biomech BioMed. Eng. Imaging Vis.* 6:3, 283–292. doi: 10.1080/21681163.2016.1149104

Xiong, B., Wang, B., Xiong, S., Lin, C., and Yuan, X. (2019). 3D morphological processing for wheat spike phenotypes using computed tomography images. *Remote Sens* 11, 1110. doi: 10.3390/rs11091110

Xu, Z., Wu, Z., and Feng, J. (2018). CFUN: Combining faster r-CNN and U-net network for efficient whole heart segmentation. *arXiv:1812.04914*. doi: 10.48550/arXiv.1812.04914

Zhao, Z., and Singer, A. (2014). Rotationally Invariant Image Representation for Viewing Direction Classification in Cryo-EM. *J Struct Biol*. 186(1), 153–166. doi: 10.1016/j.jsb.2014.03.003

Zhao, Z., Yang, L., Zheng, H., Guldner, I. H., Zhang, S., and Chen, D. Z. (2018). "Deep learning based instance segmentation in 3D biomedical images using weak annotation," in *Medical image computing and computer assisted intervention (MICCAI), 2018. lecture notes in computer science*, vol. vol. 11073 . Eds. A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López and G. Fichtinger (Cham: Springer).

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 March 2023
DOI 10.3389/fpls.2023.1132909

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Shanwen Zhang,
Xijing University, China
Yunchao Tang,
Zhongkai University of Agriculture and Engineering, China

*CORRESPONDENCE
Jun Li
autojunli@scau.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 28 December 2022
ACCEPTED 17 February 2023
PUBLISHED 06 March 2023

CITATION
Li D, Sun X, Jia Y, Yao Z, Lin P, Chen Y, Zhou H, Zhou Z, Wu K, Shi L and Li J (2023) A longan yield estimation approach based on UAV images and deep learning. *Front. Plant Sci.* 14:1132909. doi: 10.3389/fpls.2023.1132909

COPYRIGHT
© 2023 Li, Sun, Jia, Yao, Lin, Chen, Zhou, Zhou, Wu, Shi and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author (s) and the copyright owner (s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A longan yield estimation approach based on UAV images and deep learning

Denghui Li[1,2†], Xiaoxuan Sun[3,4,5†], Yuhang Jia[1], Zhongwei Yao[1], Peiyi Lin[1], Yingyi Chen[1], Haobo Zhou[1], Zhengqi Zhou[1], Kaixuan Wu[1], Linlin Shi[1] and Jun Li[1,2*]

[1]College of Engineering, South China Agricultural University, Guangzhou, China, [2]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China, [3]Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [4]South China National Botanical Garden, Guangzhou, China, [5]University of Chinese Academy of Sciences, Beijing, China

Longan yield estimation is an important practice before longan harvests. Statistical longan yield data can provide an important reference for market pricing and improving harvest efficiency and can directly determine the economic benefits of longan orchards. At present, the statistical work concerning longan yields requires high labor costs. Aiming at the task of longan yield estimation, combined with deep learning and regression analysis technology, this study proposed a method to calculate longan yield in complex natural environment. First, a UAV was used to collect video images of a longan canopy at the mature stage. Second, the CF-YD model and SF-YD model were constructed to identify Cluster_Fruits and Single_Fruits, respectively, realizing the task of automatically identifying the number of targets directly from images. Finally, according to the sample data collected from real orchards, a regression analysis was carried out on the target quantity detected by the model and the real target quantity, and estimation models were constructed for determining the Cluster_Fruits on a single longan tree and the Single_Fruits on a single Cluster_Fruit. Then, an error analysis was conducted on the data obtained from the manual counting process and the estimation model, and the average error rate regarding the number of Cluster_Fruits was 2.66%, while the average error rate regarding the number of Single_Fruits was 2.99%. The results show that the method proposed in this paper is effective at estimating longan yields and can provide guidance for improving the efficiency of longan fruit harvests.

KEYWORDS

yield estimation, UAV image, convolutional neural network, image analysis, regression analysis

1 Introduction

Smart orchard systems can effectively evaluate the growth conditions of fruit trees and improve the quality of fruits through digital technology. During the fruit ripening period, accurate statistics regarding the output of each fruit tree and the total output of the whole orchard can not only improve the efficiency of deploying harvesting robots and transportation robots but also guide market pricing and upgrade the fruit yield grade, which is conducive to the maximization of the economic benefits of orchards (He et al., 2022). Longan is widely studied in the field of smart orchards. At present, the yield estimation methods for longan orchards mainly adopt manual visual investigation. This statistical method is labor-intensive, laborious and time-consuming; is easily influenced by the subjective factors of different investigators; and has low accuracy and efficiency (Marani et al., 2021). Therefore, to reduce the cost of longan orchard yield estimation and improve the accuracy of yield estimation, it is necessary to develop a system that can automatically estimate longan orchard yields.

Longan fruits are usually clustered and grow on the outside of the canopy (Pham et al., 2015). The growth characteristics of fruits and their quantitative statistical scheme are shown in Figure 1. In the natural environment, the distribution of longan fruits is complex, and they are easily blocked by leaves and branches, exhibiting different postures in different growing environments. Therefore, accurate longan fruit detection is the difficult part of realizing automatic yield estimation, which directly affects the accuracy and efficiency of orchard yield estimation. In recent years, researchers in related fields have used shape matching, color space transformation, threshold segmentation, multiscale feature fusion, fuzzy clustering and other methods to identify, detect and classify apples, oranges and other fruits (Jaisin et al., 2013; Xiong et al., 2018; Zhuang et al., 2019; He et al., 2020; Lin et al., 2020). The traditional machine learning methods used in these studies can only be used for image processing tasks with simple background conditions, and have poor robustness in the face of very complex actual orchard scenes.

With the development of sensor and computer technology, deep learning approaches have been widely developed and applied by researchers, and deep learning exhibits an excellent learning ability in cases involving the extraction of features from complex images. In recent years, with the demand for intelligence in the agricultural field, an increasing number of researchers have used deep learning technology to process collected image data for various tasks (Wang et al., 2019; da Silva et al., 2021; de Medeiros et al., 2021; Zhou et al., 2022), including fruit recognition (Gao et al., 2020; Xiong et al., 2020), classification of plants (Flores et al., 2021), classification of pests and diseases (Anagnostis et al., 2021; Singh et al., 2021), monitoring of crop growth state based on remote sensing (Ma et al., 2019; Paoletti et al., 2019), nondestructive testing and grading of fruit (Koirala et al., 2019), and animal behavior analysis (Norouzzadeh et al., 2018). To sum up, the deep learning model has stronger feature extraction ability, and it can effectively solve complex nonlinear problems. Faced with the problems of complex image backgrounds, uneven light intensities and diverse fruit features in complex orchard scenes, some researchers have applied deep learning technology to target detection tasks in complex scenes, and these models have strong robustness (Alpaydin, 2016; Liang et al., 2020; Li et al., 2021; Zhong et al., 2021; Wu et al., 2022; Tang et al., 2023).

How to quickly obtain high-definition images of orchard scenes is a key issue for improving the efficiency of orchard yield estimation. With the rapid development of unmanned aerial vehicle (UAV) power systems, control systems and sensor technology, it is possible for UAVs to carry various types of sensors to observe the earth. In recent years, UAVs have been equipped with various sensors and used in agriculture, including plant growth state detection and yield estimation (Vanegas et al., 2018; Tetila et al., 2020; Zhou et al., 2020; Feng et al., 2020a; Feng et al., 2020b; Sumesh et al., 2021). Therefore, the researchers on our team use an RGB camera on a UAV to plan the UAV flight path in advance and quickly obtain high-definition image data about the orchard.

In this study, by combining a UAV and deep learning application technology, a fast and accurate longan yield statistics approach is proposed. This approach will help to improve the accuracy and efficiency of the yield statistics of each longan tree in the modern orchard production scene and provide information for the task assignments of fruit picking UAVs and fruit transport aircraft. The main contributions of this research are as follows.

a. A method of collecting canopy images and videos of each fruit tree with a UAV is proposed to accurately and completely obtain canopy image data for each fruit tree.
b. Two datasets are set up to train and evaluate the performance of different target detection models.
c. A scheme for counting the numbers of different targets is proposed; this scheme includes a model based on Cluster_Fruit-YOLOv5s_Deepsort (CF-YD) and a model based on Single_Fruit-YOLOv7_Deepsort (SF-YD).
d. A regression analysis is carried out on the quantities counted by the two models and their real quantities to obtain a fitting equation.

The main contents of this paper are as follows: Section 2 introduces the materials and methods, Section 3 introduces the model construction process and the statistical strategy for

Abbreviations: UAV, Unmanned aerial vehicle; CNN, Convolutional neural network; YOLO, You Only Look Once; SORT, Simple Online and Realtime Tracking; CF-YD, Cluster_Fruit-YOLOv5s_Deepsort; SF-YD, Single_Fruit-YOLOv7_Deepsort; RGB, Red Green Blue; ID, Identity document; IoU, Intersection over Union; R-CNN, Recursive convolutional neural network; SPP, Spatial Pyramid Pooling; FPN, Feature pyramid network; PAN, Path aggregation network; NMS, Non Maximum Suppression; E-ELAN, Extended Efficient Long-Range Attention Network; PAFPN, Path Aggregation Feature Pyramid Network; P, Precision; R, Recall; AP, Average precision of a category; FPS, Frames per second; TP, True positive; FP, False positive; TN, True negative; FN, False negative; P-R, Precision-Recall; IDS, Identity Switch; MOTA, Multiple Object Tracking Accuracy; MOTP, Multiple Object Tracking Precision.

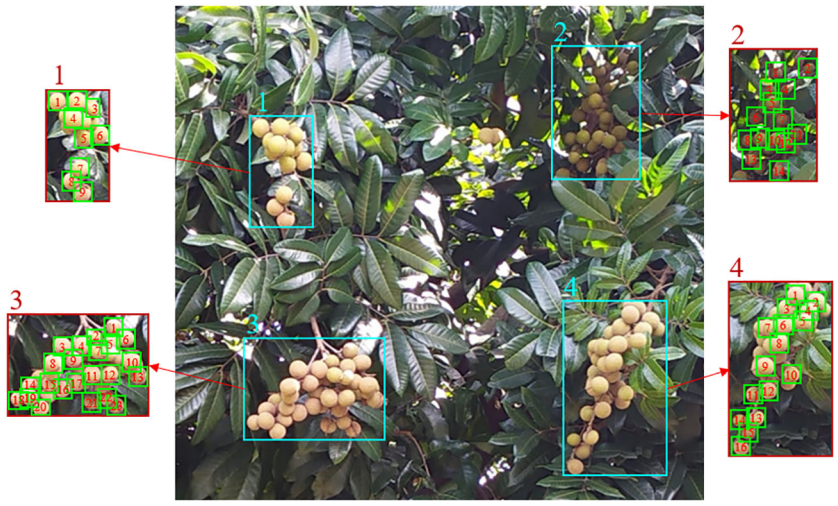


FIGURE 1
Growth characteristics of longan fruits and their quantitative statistical scheme.

calculating single fruit tree yields, Section 4 introduces the model experiment and results analysis in detail, and Section 5 summarizes the full text.

2 Materials and methods

2.1 Overview of the fruit tree yield estimation methods

Figure 2 shows the method of quickly and accurately calculate the yield of each fruit tree. First, fruit tree canopy images are collected, and the obtained images are preprocessed as the Cluster_Fruit image dataset. Second, a two-step model is established. The first step is to count the number of Cluster_Fruits on a single fruit tree based on the CF-YD model, and the second step is to count the number of Single_Fruits on each Cluster_Fruit based on the SF-YD model. The results of each step are combined with the corresponding fitting equation to correct the final result.

In the two-step model of this scheme, when each Cluster_Fruit is detected from a canopy image of fruit trees based on the CF-YD model, each Cluster_Fruit needs to be cut out as the input image of the statistical Single_Fruit quantity model. After using the SF-YD model to count the number of Single_Fruits in each image cut out in the previous step, it is necessary to calculate the yield of a single fruit tree according to the weight of the Single_Fruit.

2.2 Sensor system and image acquisition

To build the data set, 1100 valid longan images and 16 complete longan tree canopy videos were acquired at the Longan orchard in Guangzhou on July 1-25, 2021, and July 1-26, 2022, during two different time periods: morning (7:30–11:30) and afternoon (13:30–18:30). Furthermore, to perform modelling and verify the accuracy of the model, the actual numbers of Cluster_Fruits and Single_Fruits and the quality of 16 Cluster_Fruits from 16 fruit trees in the orchard were manually counted. A lightweight, small-size and high-resolution RGB camera mounted on a DJI Mavic 2 Pro UAV was used to collect orchard canopy images. The camera had 12 million pixels, the viewing angle was 85 degrees, the focal point ranged from 0.5 m to infinity, and 120 images could be taken in one second at the fastest speed. Figure 3 is a schematic diagram of the image acquisition mode.

According to the planting mode of modern orchards, the images were collected according to the following steps. ① Control the UAV to fly around the fruit trees, and during this process, have the RGB camera mounted on the UAV always look straight at the tree center to collect the canopy images of the fruit trees. ② Set the RGB image resolution to 1280×720 pixels, and automatically save each image to the image acquisition card obtaining it. In order to ensure the diversity of images in the data set, images were taken on sunny and cloudy days respectively, including images of Shixia longan and Chuliang longan. Figure 4 shows the examples of the UAV images.

2.3 Image preprocessing

In this study, data sets are prepared for the two-step model. First, aiming at the images collected by the camera on the UAV, image cropping and size normalization are performed, and the bounding box of each Cluster_Fruit in each sample is manually marked, forming a Cluster_Fruit data set for the CF-YD model. Second, the trained CF-YD model is used to detect Cluster_Fruits in the original RGB images, each Cluster_Fruit image is cut out according to the output coordinate information, and the

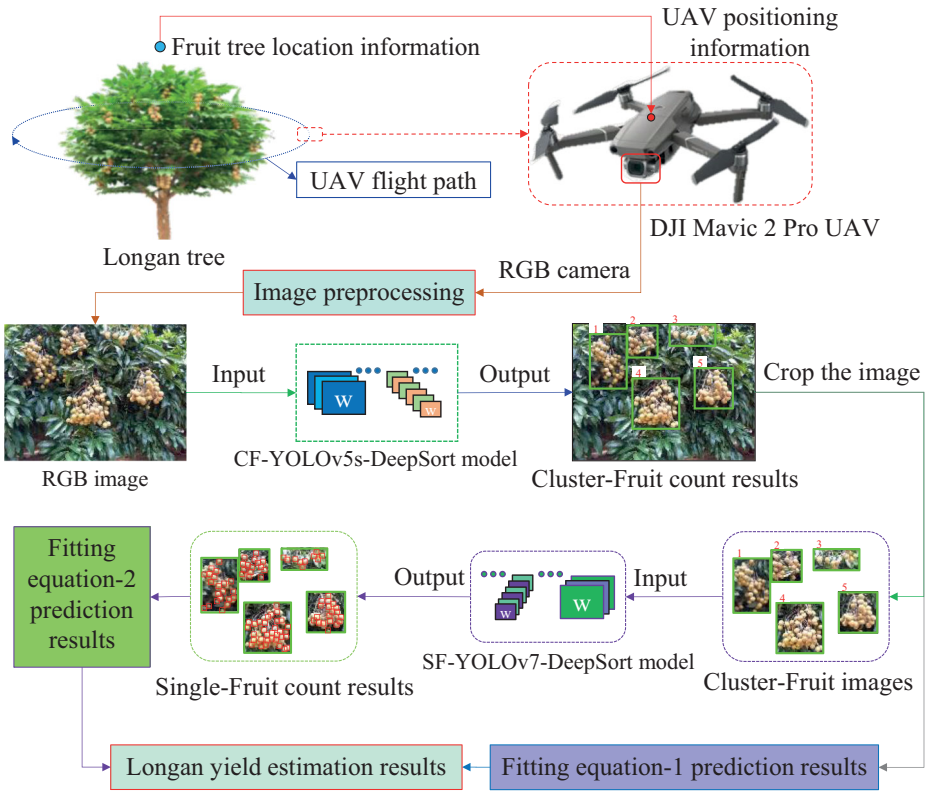


FIGURE 2
Solution for longan fruit tree yield statistics.

Single_Fruit positions in each image are manually marked to form a Single_Fruit data set for the SF-YD model. Figure 5 shows the whole process.

2.3.1 Construction of the Cluster_Fruit image data set

During the process of constructing the Cluster_Fruit image dataset, firstly, to ensure the diversity of the training samples, a self-programmed random cropping algorithm is used to amplify the Cluster_Fruit images. After the initial longan dataset is expanded, 1100 longan images are obtained. Then, the image size is normalized to 1280×1280 pixels. Finally, the 1100 images are manually annotated following the guidelines of the Pascal VOC 2010 reference challenge. The labelling information mainly includes the size of each image, the target category and the specific position coordinate information of the target area.

2.3.2 Construction of the Single_Fruit image dataset

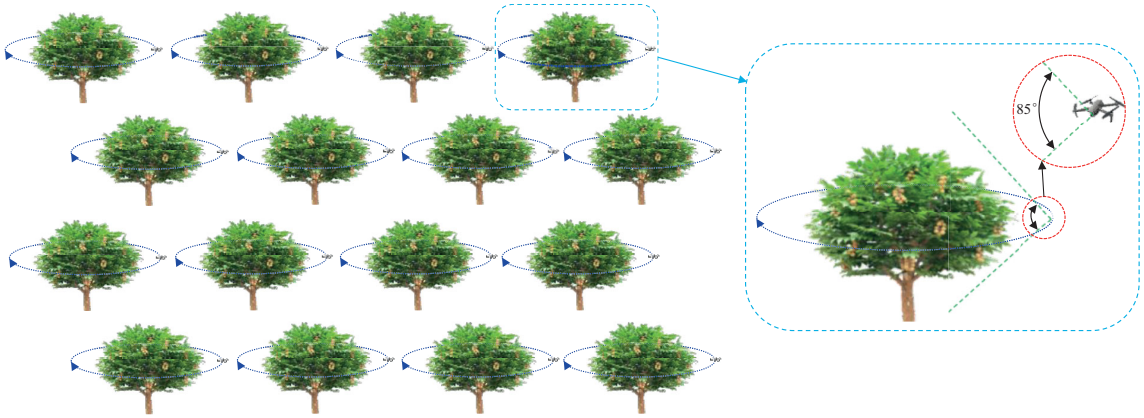
As the SF-YD model adopted in this study first normalizes the input images of any size to 640×640 pixels, the trained CF-YD model is used to detect each image in the Cluster_Fruit data set, and the coordinate information of each Cluster_Fruit image is obtained and directly cut out from the original image. After screening, 1100 Cluster_Fruit images are selected. LableImg software is used to manually mark the positions of Single_Fruits in each image to form a Single_Fruit data set for training and testing the SF-YD model.

In summary, the Cluster_Fruit data set and Single_Fruit data set are constructed by the above two methods. The images in both data sets are divided into a training set, verification set, and test set according to a ratio of 8:1:1. Table 1 presents the images and the annotation information.

3 Model construction and statistical strategy for fruit tree yield estimation

In this section, according to the video image data of the canopy of a single fruit tree collected by the RGB camera on the UAV, a yield prediction scheme for a single fruit tree is proposed. First, according to the growth characteristics of Cluster_Fruit, the improved YOLOv5s target detection algorithm and Deepsort target tracking algorithm are incorporated into the CF-YD model, and the flow chart for quickly and accurately obtaining the numerical and location information of Cluster_Fruits from the video image of a single fruit tree is determined. Then, according to the growth characteristics of Single_Fruits, the YOLOv7 target detection algorithm and Deepsort target tracking algorithm are merged into the SF-YD model, and the flow chart for quickly obtaining the number of Single_Fruits from a video image of

10.3389/fpls.2023.1132909



A (Schematic diagram of the flight path when the UAV acquired the images) **B** (Camera view of the UAV)

FIGURE 3
Schematic diagram of the image acquisition mode.

Cluster_Fruit is worked out. Finally, according to the prediction results regarding the quantities of Cluster_Fruits and Single_Fruits, a strategy for counting the output of a single fruit tree is proposed.

3.1 Deepsort algorithm

The Deepsort algorithm (Wojke et al., 2017) is an algorithm with a multitarget tracking function, which is improved on the basis of the SORT algorithm (Bewley et al., 2016). Compared with the SORT algorithm, the Deepsort algorithm improves the content matching process to avoid ignoring multitarget ID transformations, uses appearance information to curb the frequency of target ID transformations, and adds a simple convolutional neural network (CNN) model to extract the appearance features of detected targets (expressed by low-dimensional vectors). The core of the Deepsort algorithm consists of prediction, observation and updating. The specific flow of the Deeppart algorithm is as follows. ① The target information predicted by You Only Look Once (YOLO) is input into the Deeppart algorithm as the observed value. The Kalman filter first judges whether a track is present, and if one is, it predicts the prior probability of the target information, then carries out



A (Sunny day, Shixia longan) **B** (Sunny day, Chuliang longan)

C (Cloudy day, Shixia longan) **D** (Cloudy day, Chuliang longan)

FIGURE 4
Examples of the UAV images.

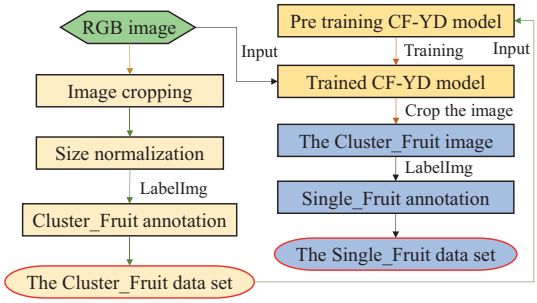


FIGURE 5
Diagram of the whole process.

cascade matching and IoU matching in the matching module, and finally obtains the matching success list. ② In the Kalman updating module, a posteriori prediction is performed on the successfully matched target to obtain the corrected target coordinates, and parameters such as the Kalman gain are updated. ③ The above operations are repeated until all the videos are processed.

3.2 CF-YD model for Cluster_Fruit statistics

The clusters of longan fruits are usually distributed in the canopies of fruit trees in disordered arrangements, and their sizes become very different as the distance from the camera increases. When quickly and accurately counting the number and locations of longan Cluster_Fruits from the canopy images of fruit trees, it is necessary to overcome the problem that target scale changes greatly affect the resulting detection accuracy. The most commonly used target detection algorithms are the R-CNN series (Girshick et al., 2015a; Girshick, 2015b; Ren et al., 2017) and YOLO series (Redmon et al., 2016; Redmon & Farhadi, 2018; Bochkovskiy et al., 2020) models. R-CNN series algorithms, also known as target detection algorithms based on candidate areas, first generate candidate areas that may contain objects and then further classify and calibrate the candidate areas to obtain the final detection results. During the training process, YOLO series algorithms can pay more attention to the global information and the whole image in target detection. The core idea of YOLO is to use the whole picture as the input of the network and directly return to the position and category of the bounding box at the output.

Compared with the classic YOLOv3 algorithm, the data enhancement step of the YOLOv5 detection algorithm uses Mosaic to expand the input dataset, and it can also perform operations such as flipping, brightness adjustment and clipping. For a sample set with less data, the data can be effectively expanded. Four versions of YOLOv5 are available, namely, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, among which YOLOv5s is the network with the smallest depth and the smallest characteristic map width in this series of detection networks.

YOLOv5s mainly consists of three parts: a backbone, a neck and an output. The backbone is the basic feature extraction layer, which is used to extract feature information from images. It includes four modules: Focus, CBH, CSP1-x and spatial pyramid pooling (SPP) modules (He et al., 2015). The neck is a feature fusion layer whose function is to fuse image information with different scales to obtain better detection results. It uses the rectified linear unit (ReLU) activation function and adopts a feature pyramid network (FPN) (Lin et al., 2017) + PAN (Liu et al., 2018) network structure. The output is the output layer, whose function is to output the predicted target information, in which nonmaximum suppression (NMS) is performed on the last detection frame of the target to obtain the optimal target frame; three different detection scales (20×20, 40×40, 80×80) are provided, which can predict longan Cluster_Fruits with different sizes. In the early stage, our team improves the YOLOv5s model for the Cluster_Fruit detection task and improves the accuracy of the model in the target detection task.

In this paper, according to the growth characteristics of Cluster_Fruit, the improved YOLOv5s target detection algorithm (Li et al., 2022) and Deepsort target tracking algorithm are incorporated into the CF-YD model. Figure 6A shows the flow chart for the numerical Cluster_Fruit statistics in a single fruit tree. First, after the complete video data of a single fruit tree are input

TABLE 1 Details of the Cluster_Fruit and Single_Fruit data sets.

Data set	Cluster_Fruit data set		Single_Fruit data set	
	Images	Cluster_Fruit	Images	Single_Fruit
Full data set	1100	11239	1100	24045
Training data set	880	8898	880	19393
Validation data set	110	1205	110	2285
Test data set	110	1136	110	2367

10.3389/fpls.2023.1132909

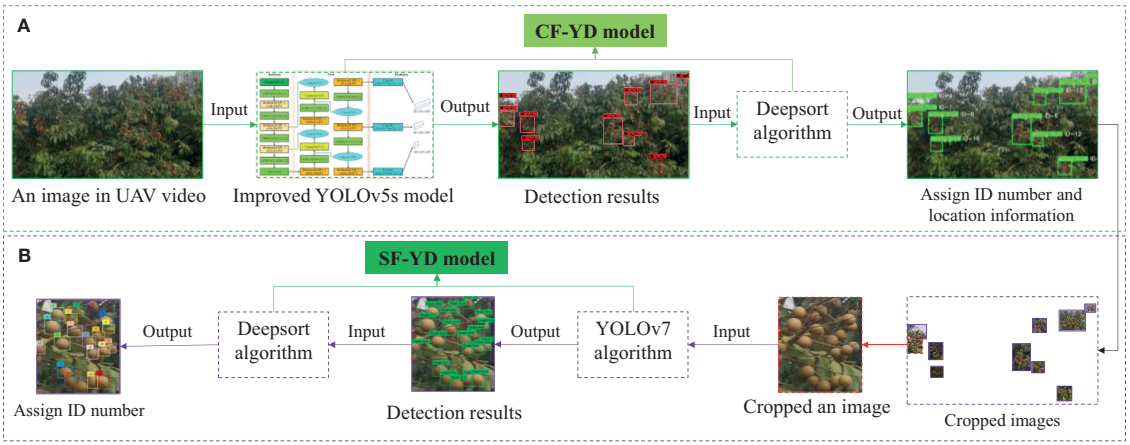


FIGURE 6
Flow chart of Cluster_Fruits and Single_Fruits quantity prediction. **(A)** The flow chart of Single_Fruit quantity statistics in a Cluster_Fruit, **(B)** The flow chart of Cluster_Fruit quantity statistics in a single fruit tree.

into the YOLOv5s algorithm, the YOLOv5s algorithm detects multiple targets in each frame and inputs the position information obtained from multiple Cluster_Fruits into the Deepsort algorithm to assign ID numbers. Then, the correlation filtering algorithm is used to compare whether anchor frames with the same size are present in the front and back frames (a target with the same anchor frame size continues to use the original number and assigns a new ID number to the new target). Finally, the maximum ID number is output, and the value of this number is used as the number of predicted Cluster_Fruits in the fruit tree canopy.

3.3 SF-YD model for Single_Fruit statistics

The Single_Fruits of longan are usually distributed in Cluster_Fruits in a disorder manner and occupy very small areas in the canopy images of fruit trees, so it is difficult to accurately count the number of Single_Fruits in each Cluster_Fruit directly from the canopies of fruit trees. In this study, the latest YOLOv7 algorithm (Wang et al., 2022), which has a higher detection accuracy and a faster detection speed than other algorithms in the series, is adopted to realize the Single_Fruit detection task. The YOLOv7 algorithm adopts strategies such as the extended efficient long-range attention network (E-ELAN), model scaling based on concatenation-based models (Wang et al., 2021), and convolutional reparameterization (Ding et al., 2021) and achieves a very good balance between detection efficiency and accuracy.

The YOLOv7 network mainly includes four parts: an input, a backbone, a head and a prediction module. The input module normalizes the input image of any size to the input pixel size set by the backbone network. The backbone module consists of several BConv layers, E-ELAN layers and MPConv layers. The BConv layer consists of a convolution layer, a batch normalization layer and a LeakyReLU activation function (Jiang & Cheng, 2019), which is used to extract image features with different scales. The E-ELAN layer keeps the original ELAN design framework and improves the learning ability of the network without destroying the original gradient path by guiding the computing blocks of different feature groups to learn more diverse features. On the basis of the BConv layer, the MPConv layer adds an Maxpool layer to form two branches. The upper branch cuts the image length and width by half through Maxpool and the image channel by half through the BConv layer. In the lower branch, the image channel is halved by the first BConv layer, and the image length and width are halved by the second BConv layer. Finally, the features extracted by the upper and lower branches are fused by the Cat operation, which improves the feature extraction ability of the network. The head module is composed of a path aggregation FPN (PAFPN) (Ge et al., 2021) structure. By introducing the bottom-up path, the bottom-up information can be transmitted to the higher level more easily, thus realizing the efficient integration of features at different levels. The prediction module adjusts the number of image channels for three features of the PAFPN output with different scale, such as P3, P4 and P5, through the REPVGG block structure and finally uses a 1×1 convolution to predict the confidence, category and anchor frame.

According to the growth characteristics of Single_Fruits, the YOLOv7 target detection algorithm and Deepsort target tracking algorithm are incorporated into the SF-YD model. Figure 6B shows the flow chart of the numerical Single_Fruit statistics in a Cluster_Fruit. First, each Cluster_Fruit with an ID number assigned by the CF-YD model is continuously cut out from the original video image to form Cluster_Fruit video data with ID numbers. The YOLOv7 algorithm detects multiple targets in each frame, inputs the obtained position information of multiple Single_Fruits into the Deepsort algorithm to assign ID numbers, then compares whether anchor frames of the same size are present in the previous and subsequent frames by using the correlation filtering algorithm (targets with the same anchor frame size

continue to use the original number, and new ID numbers are assigned to new targets), and finally outputs the maximum ID number and takes the value of this number as the number of Single_Fruits in each Cluster_Fruit.

3.4 Statistical strategy for a single fruit tree yield

In this section, according to the prediction results regarding the numbers of Cluster_Fruits and Single_Fruits in the previous two sections, the specific steps of longan yield estimation are formulated:

(1) Step 1: The CF-YD model is used to quickly obtain the number N_{CF1} of Cluster_Fruits from the canopy images of fruit trees, and the total number N_{CF2} of Cluster_Fruits is predicted by establishing a regression analysis model with the number of Cluster_Fruits counted in the real orchard.

(2) Step 2: The location information of the N_{CF1} Cluster_Fruits obtained in the previous step is cut from the original image in turn and input into the SF-YD model to obtain the total number N_{SF1} of Single_Fruits. The total number N_{SF2} of Single_Fruits is predicted by establishing a regression analysis model with the number of Single_Fruits in the N_{CF1} Cluster_Fruit statistics in the actual orchard.

(3) Step 3: According to the total number N_{SF2} of Single_Fruits in N_{CF1} Cluster_Fruits, the average number AVE_{nSF} of Single_Fruits in a single cluster can be calculated. Ten Cluster_Fruits are randomly selected to weigh and count the number of Single_Fruits and calculate the average quality AVE_{mSF} of Single_Fruits.

(4) Step 4: The formula for calculating the yield TQ of a single fruit tree is:

$$TQ = (N_{SF2} + (N_{CF2} - N_{CF1}) * AVE_{nSF}) * AVE_{mSF} \quad (1)$$

4 Model experiment and results analysis

4.1 Model training and parameter design

The training and testing processes of the CF-YD and SF-YD models are implemented on a workstation with the Ubuntu 18.04 LTS operating system. The main hardware devices of the workstation are as follows: GPU: NVIDIA GTX3060 (configured with CUDA 10.1 and cuDNN 7.1); processor: 11th Gen Intel (R) Core (TM) i7-11800H; RAM: NVIDIA 16G; and hard disk: Samsung 1T. On the PyTorch deep learning framework, a CNN model is built with the Python programming language.

The image size of the CF-YD model training data is set to 1280×1280 pixels. In terms of parameter settings, the intersection over union (IoU) is set to 0.5, the initial learning rate is 1e-4, and the learning rate at the end of training is set to 1e-5. The dataset is split at a 90-10 training-verification ratio, and a total of 500 epochs of iterative training are conducted. During the training process, after conducting a series of convolution and pooling operations with the CF-YD model, the input image uses the anchor box in the feature map layer to extract a series of features. In the feature layer, each cell is mapped to the original image, the premarked anchor box is found, and then the loss value between this anchor box and the ground truth is calculated. After training, the CF-YD model obtains a series of model parameters to fit the real border with the anchor box.

During the training process of the SF-YD model, input images of any size are normalized to 640×640 pixels through the input module. The training process of the model is divided into the Freeze phase and UnFreeze phase, the optimizer is set to stochastic gradient descent (SGD), the initial learning rate is 0.01, the momentum is 0.9, and the weight decay is set to 5e-4. The cosine annealing algorithm is used to adjust the learning rate, and the minimum learning rate is 1e-4. The training and verification steps alternate. During the Freeze phase, the training duration is 50 epochs, each epoch has 220 iterations in the training phase and 28 iterations in the verification phase, and the batch size is 4. During the UnFreeze phase, the training duration is 250 epochs, with 440 iterations in the training phase and 55 iterations in the verification phase of each epoch, and the batch size is 2. After training, the SF-YD model obtains a series of model parameters to fit the real border with the anchor box.

During the training and testing processes of the CF-YD and SF-YD models, it is necessary to generate a series of anchor boxes (candidate areas) in the given image according to certain rules. In this study, k-means clustering and a genetic algorithm are used to obtain anchor boxes. Because the prediction layer of the YOLO network contains three scales of information (corresponding to three receptive fields), each scale contains three anchors. Therefore, the YOLO network needs nine anchor scales; that is, the sizes of all the target bounding boxes in the dataset are clustered into nine categories. Through the analysis of the Cluster_Fruit and Single_Fruit datasets, the k-means clustering results of all the target bounding boxes in the two datasets are obtained. Each point in Figure 7A corresponds to a target bounding box in the Cluster_Fruit dataset. According to the overall size characteristics of the target bounding boxes, nine types of anchor boxes that are suitable for training and testing the CF-YD model are determined as [16,16, 21,28, 28,23, 30,39, 41,33, 46,52, 67,77, 116,135, 247,291]. Each point in Figure 7B corresponds to a target bounding box in the Single_Fruit dataset. According to the overall size characteristics of the target bounding box, nine types of anchor boxes that are suitable for training and testing the SF-YD model are determined as [45,43, 61,36, 61,53, 78,44, 70,66, 93,54, 87,78, 131,74, 110,104]. At the same time, it can be seen from the figure that the more points there are with the same color, the more targets with this cluster size, and the points with different colors represent targets with different cluster sizes. In other words, this figure can reflect the complicated situation regarding the targets to be detected in an orchard scene to some extent.

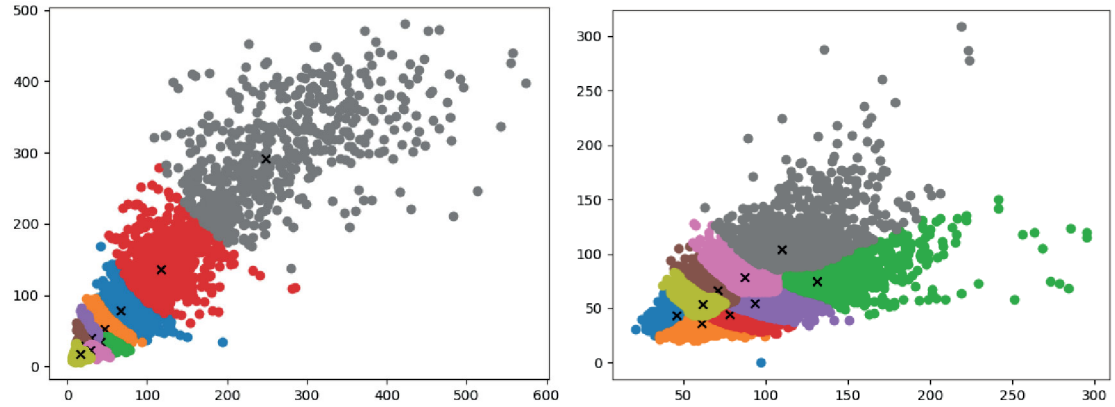


FIGURE 7
Clustering results of the target bounding boxes in the two datasets.

4.2 Model evaluation indicators

4.2.1 Evaluation indices for the object detection algorithm

In this study, P, R, F1 score, AP, and FPS were used to evaluate the performance of two target detection models. The calculation methods for calculating the P, R, F1 score and AP here are shown in formulas (2), (3), (4) and (5).

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1 score} = \frac{2 \star P \star R}{P + R} \tag{4}$$

$$AP = \int_0^1 P(R) d_R \tag{5}$$

In these formulas, TP represents true cases, FP represents false-positive cases, TN represents true-negative cases, and FN represents false-negative cases.

4.2.2 Evaluation indices for the Deepsort algorithm

This study selects identity switches (IDSs), multiple-object tracking accuracy (MOTA) and multiple-object tracking precision (MOTP) to evaluate the effectiveness of the multitarget tracking algorithm. IDS is the number of times the tracking target ID changes. The smaller its value is, the better the tracking stability. MOTA considers false alarms and IDSs simultaneously and measures the performance of the tracking algorithm in terms of detecting targets and keeping track of them, which has nothing to do with target detection accuracy. The larger its value is, the better the performance of the algorithm. MOTP is used to quantify the positioning accuracy of the detector. The larger its value is, the higher the accuracy of the detector.

4.3 Results and discussion of the object detection and counting tasks

4.3.1 Performance evaluation results of different models

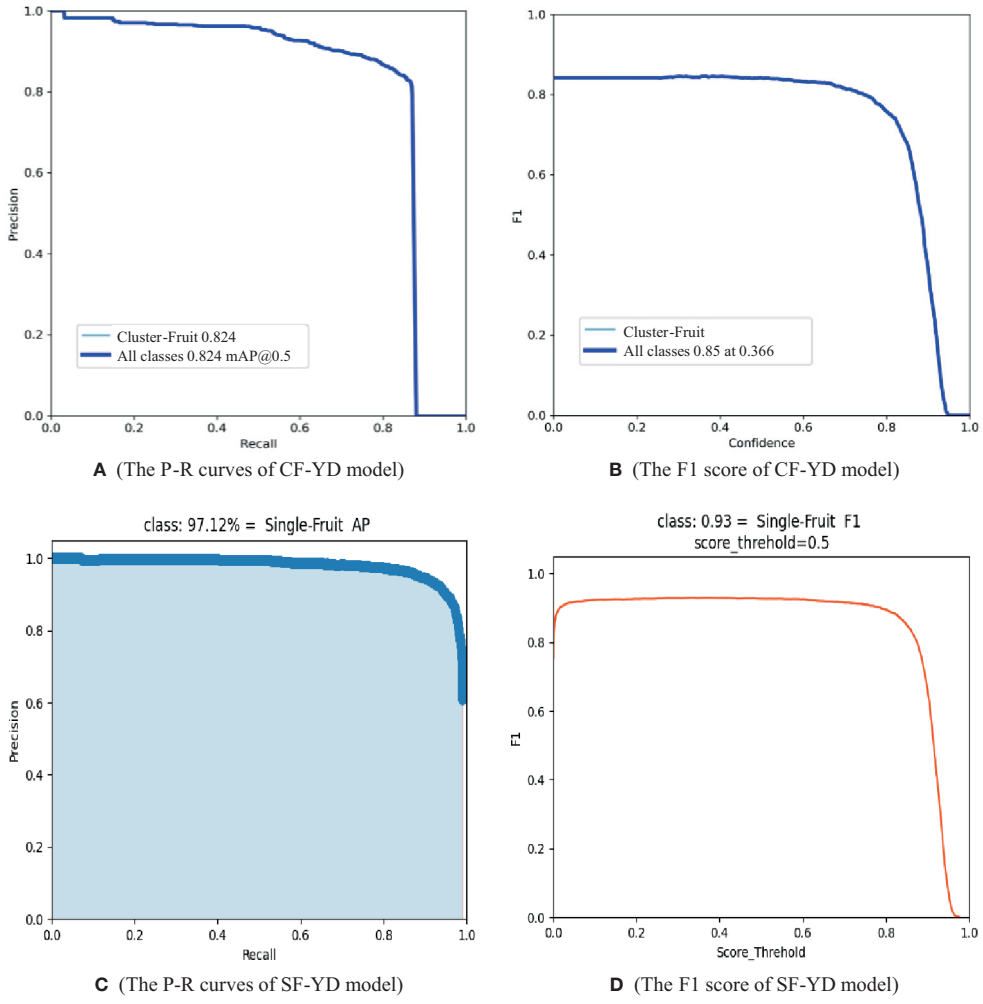
To fully evaluate the performance of the CF-YD model in detecting Cluster_Fruits and the SF-YD model in detecting Single_Fruits, first, the CF-YD and SF-YD models are trained according to the training parameters set in Section 4.1, and the weight file with the best training effect in each model is used as the weight file for testing the model performance. Then, the CF-YD model for the Cluster_Fruit test set and the SF-YD model for the Single_Fruit test set comprehensively evaluated from the aspects of P, R, AP, FPS, F1 score, etc., and the obtained results are shown in Table 2.

The P-R curve and F1 score changes exhibited by the CF-YD model on the Cluster_Fruit test dataset are shown in Figures 8A, B,

TABLE 2 Evaluation index results obtained on the test dataset under different models.

Models/Evaluation index	P (%)	R (%)	AP (%)	F1 score	FPS
CF-YD model (Cluster_Fruit test data set)	83.50	85.70	82.40	0.85	56.21
SF-YD model (Single_Fruit test data set)	93.23	91.97	97.12	0.93	98.35

10.3389/fpls.2023.1132909



A (The P-R curves of CF-YD model)

B (The F1 score of CF-YD model)

C (The P-R curves of SF-YD model)

D (The F1 score of SF-YD model)

FIGURE 8
P-R curves and F1 scores of different detection methods.

respectively. The area enclosed by the P-R curve and the two coordinate axes in Figure 8A corresponds to the AP value of Cluster_Fruit detection. As shown in Table 2, the AP value of the CF-YD model for Cluster_Fruit detection is 82.4% on the test set, and the detection accuracy is high. However, some Cluster_Fruits are still blocked by other Cluster_Fruits or branches and leaves and cannot be accurately detected. The F1 score in Figure 8B first changes slightly with increasing confidence value and then suddenly decreases sharply when the confidence value is greater than 0.7. Therefore, it is usually sufficient to set this parameter to 0.5 in the model training stage. The FPS value is the number of images that the model can detect per second, and the detection time of each image is only 18 ms. According to the clustering results of the target sizes in the Cluster_Fruit dataset, the Cluster_Fruit size exhibits the diversity characteristic. The above results show that the CF-YD model has good detection performance for multiscale targets.

The changes in the P-R curve and F1 score of the SF-YD model on the Single_Fruit test dataset are shown in Figures 8C, D, respectively. The area enclosed by the P-R curve and the two coordinate axes in Figure 8C basically covers the whole coordinate system. As shown in Table 2, the AP value of the SF-YD model on the test set for Single_Fruit is 97.12%, demonstrating high detection accuracy, and only a few Single_Fruits are undetected. The F1 score changes slightly with increasing confidence and suddenly decreases sharply when the confidence is greater than 0.85. Therefore, it is usually sufficient to set this parameter to 0.5 in the model training stage. The FPS value is the number of images that the model can detect per second, and the detection time of each image is only 10 ms. According to the clustering results of the target sizes in the Single_Fruit dataset, the sizes of Single_Fruits are generally small. The above results show that the SF-YD model also has good detection performance for small targets.

4.3.2 Detection effects of different models in real scenes

Many varieties of longan are available, and new varieties have appeared in recent years. To further evaluate the performance of the CF-YD and SF-YD models in detecting longan Cluster_Fruits and Single_Fruits in real and complicated mountain orchard

environments, this section selects images of longan orchards with different varieties (Chuliang longan and Shixia longan), different illumination conditions (Sunny day and Cloudy day), different scales and different densities, tests the trained CF-YD and SF-YD models, and obtains the detection results of each model.

Figures 9A, B are the test results obtained by the CF-YD model for Chuliang longan in different scenes of real orchards. Figures 9C, D are the test results obtained by the CF-YD model for Shixia longan in different scenes of real orchards. From the detection results, it can be seen that regardless of the longan variety and in sunny day or cloudy day, Cluster_Fruit is accurately detected for large-scale or small-scale targets. The above detection results show that the CF-YD model has good feature extraction performance, has strong generalization to different varieties of longan in real orchard environments and is not easily disturbed by uneven light. It also has a good detection effect on small targets, so it is suitable for target detection in longan orchards.

To evaluate the detection effect of the SF-YD model on different longan varieties in a real orchard scene, the CF-YD model is first used to identify Cluster_Fruit in a real orchard scene and cut out the video data to form Single_Fruits, which are then input into the SF-YD model for Single_Fruit detection. Figures 10A, B are the test results obtained by the SF-YD model for Chuliang longan and Shixia longan, respectively, in different scenes of real orchards. The fruit colors and shapes of the two longan species are quite different. They exhibit different glosses at different distances and under different light. It can be seen from the detection results that Single_Fruits of different varieties are accurately detected in, different weather conditions, with different scales and in scenes with different densities. The above detection results show that the SF-YD model has good feature extraction performance, strong generalization for different varieties of longan Single_Fruits in a real orchard environment, and a good detection effect for multiscale targets, so it is suitable for small target detection in longan orchards.

4.3.3 Counting results of different models in real scenes

To evaluate the tracking performance of the CF-YD model on Cluster_Fruit and the SF-YD model on Single_Fruit, a video image of a fruit tree canopy is randomly selected to test the CF-YD model, and a video image of a Cluster_Fruit is selected to test the SF-YD model. The models are comprehensively evaluated in terms of the IDS, MOTA, MOTP and other metrics, and the results obtained are shown in Table 3.

Regarding the IDS metric, the numbers of target ID changes observed during the process of tracking the target in the video images with the two models are very small at 5 and 2, respectively. The MOTA and MOTP values of the two models are basically above 90%, which shows that both tracking algorithms can track targets stably and accurately.

To further verify the performance of the CF-YD and SF-YD models in counting the numbers of longan Cluster_Fruits and Single_Fruits in the real and complicated mountain orchard environment, this section selects images of longan orchards with different varieties (Chuliang longan and Shixia longan) and different lighting scenes, tests the trained CF-YD and SF-YD models, and obtains the counting results of each model.

Figures 11A, B are the Cluster_Fruit counting results obtained in different scenes of real orchards by the CF-YD model for Chuliang longan. Figures 11C, D show the counting results



FIGURE 9
Cluster_Fruit detection results of the CF-YD model under different scenes.

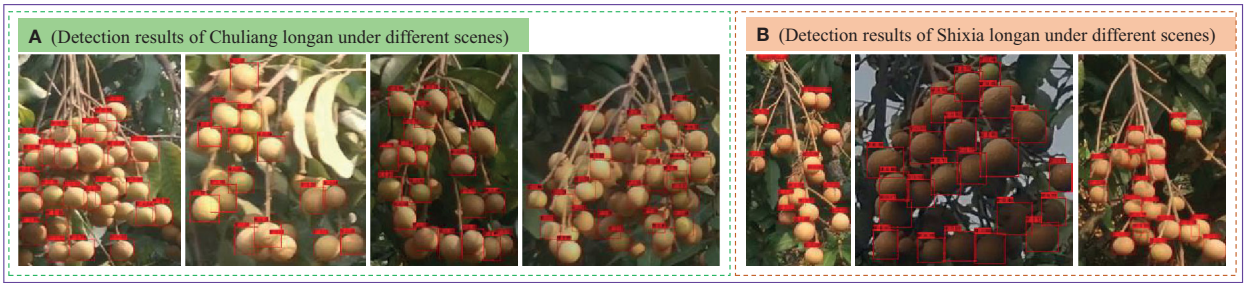


FIGURE 10
Single_Fruit detection results of the SF-YD model under different scenes.

obtained by the CF-YD model for Shixia longan in different scenes of real orchards. It can be seen from the counting results that regardless of the variety and in sunny or cloudy weather, Cluster_Fruits yield accurate counting results for large-scale or small-scale targets. The above results show that the CF-YD model has good target tracking performance. It has strong generalization for different varieties of longan in real orchard environments and is not easily disturbed by uneven lighting. It also has a good tracking effect for multiple targets, so it is suitable for target counting tasks in longan orchards.

To verify that the SF-YD model can count different longan varieties in real orchard scenes, the Single_Fruit video data are input into the SF-YD model to count Single_Fruits. Figures 12A, B are the counting results of the SF-YD model for Chuliang longan and Shixia longan, respectively, in different scenes of real orchards. It can be seen from the counting results diagram that the different varieties of Single_Fruits are accurately counted in different weather conditions. The above detection results show that the SF-YD model has a good target tracking performance, strong generalization for different varieties of longan Single_Fruit in a real orchard environment, and a good tracking effect for multiscale targets, so it is suitable for counting small targets in longan orchards.

4.4 The models for estimating the numbers of Cluster_Fruits and Single_Fruits

To accurately obtain the yield of a single longan tree, it is necessary to modify the numbers of Cluster_Fruits and Single_Fruits identified by the two models. First, 10 longan trees of different ages are randomly selected from the longan orchard, and the true number of Cluster_Fruits on each longan tree and the true numbers of Single_Fruits on the randomly selected 10 Cluster_Fruits fruits are manually counted. Then, the canopy video images of these 10 longan trees are captured by UAVs, and the number of Cluster_Fruits on each longan tree and the numbers of Single_Fruits on the 10 randomly selected Cluster_Fruits are identified by the method described in the previous section. Finally, the number of artificial statistics and the number identified by the model are fitted by an equation, and a number estimation model for the Cluster_Fruits on a single longan tree and a number estimation model for the Single_Fruits on a single Cluster_Fruit are constructed.

TABLE 3 Evaluation index results obtained by different models.

Models/Evaluation index	IDS	MOTA (%)	MOTP (%)
CF-YD model (Cluster_Fruit)	5	95.30	92.60
SF-YD model (Single_Fruit)	2	97.20	94.70

Table 4 counts the quantity information of the manual counting approach and two identification models. The actual value of Cluster_Fruits on 10 longan trees ranges from 91 to 312, and the actual value of Single_Fruits on ten Cluster_Fruits ranges from 18 to 32. Because the 10 longan trees and 10 Cluster_Fruits are randomly selected, the numbers of fruits will be different in different runs. At the same time, during the process of growth, the fruit of longan trees is affected by external conditions such as nutritional components and light conditions, so the yield of each tree is different. Exponential fitting, linear fitting, logarithmic fitting, binomial fitting, power fitting, etc., are performed for determining the numbers identified by the models and the actual number of manual statistics in Table 4. After performing a comprehensive analysis and comparing the fitting results, as shown in Figure 13, the best fitting method for the number of Cluster_Fruits on a single fruit tree is binomial fitting. The fitting equation is $y = 0.0023x^2+0.7155x+19.562$, and the determination coefficient R^2 is 0.9970. The best fitting method for the number of Single_Fruits on a single cluster is exponential fitting, the fitting equation is $y = 7.822e^{0.0565x}$, and the determination coefficient R^2 is 0.9953. Strong correlation is observed between the two samples.

4.5 Experimental results of Cluster_Fruits and Single_Fruits in real orchard scenes

To further verify the quantity estimation model in Section 4.4, six other longan trees are randomly selected from real orchards, and the true number of Cluster_Fruits on each longan tree and the true numbers of Single_Fruits on the 6 randomly selected Cluster_Fruits are obtained by manual counting. Then, the CF-YD and SF-YD models are used to obtain the identification numbers of the Cluster_Fruits and Single_Fruits from the video data, respectively. By using the fitting equation obtained in Section 4.4, the identified

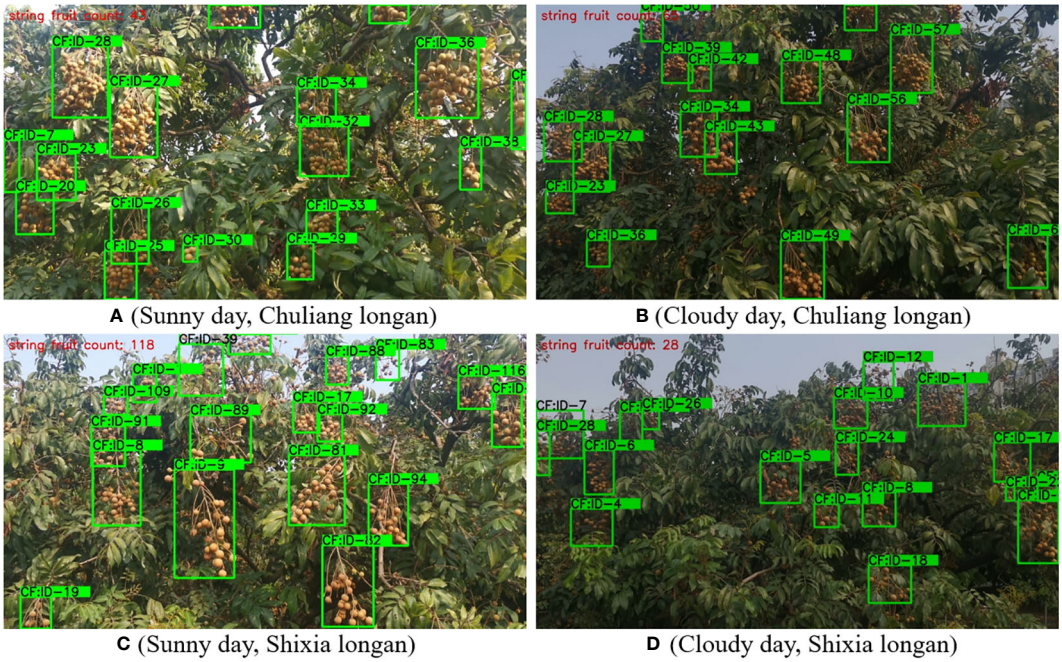


FIGURE 11
Cluster_Fruit counting results of the CF-YD model under different scenes.

numbers are corrected, and the predicted numbers of Cluster_Fruits and Single_Fruits are obtained. Finally, the error between the real quantity and the predicted quantity is analyzed. The error in this study is the absolute value of the predicted value minus the actual value, and the error rate is equal to the percentage value obtained by dividing this error by the actual value. The calculation formula for the error rate is:

$$\text{Error rate} = \left| \frac{\text{Predicted value} - \text{Actual value}}{\text{Actual value}} \right| \times 100\,\% \qquad (6)$$

The actual numbers, identified numbers and predicted numbers of Cluster_Fruits on six longan trees and Single_Fruits on 6 Cluster_Fruits are counted in Figures 14A, B, respectively, and their error rate data are counted in Figures 14C, D. It can be seen from the data in Figures 14A, B that the number of Cluster_Fruits identified by the CF-YD model and the number of Single_Fruits identified by the SF-YD model are corrected by the fitting equation obtained in Section 4.4, and the predicted numbers are very close to the actual numbers. According to the data in Figure 14C, the average error rate of Cluster_Fruit of 6 longan trees is 2.66%. According to the data in Figure 14D, the average error rate for the Single_Fruits of 6 Cluster_Fruits is 2.99%. It can be seen from the data in Figures 14C, D that the prediction error rates of Cluster_Fruits and Single_Fruits are below 5%.

According to the statistical results of Section 4.3, Section 4.4 and this section, there are errors between the real value and the identified value, which are mainly caused by two reasons. ① The two models have certain accuracy levels when detecting targets. ②

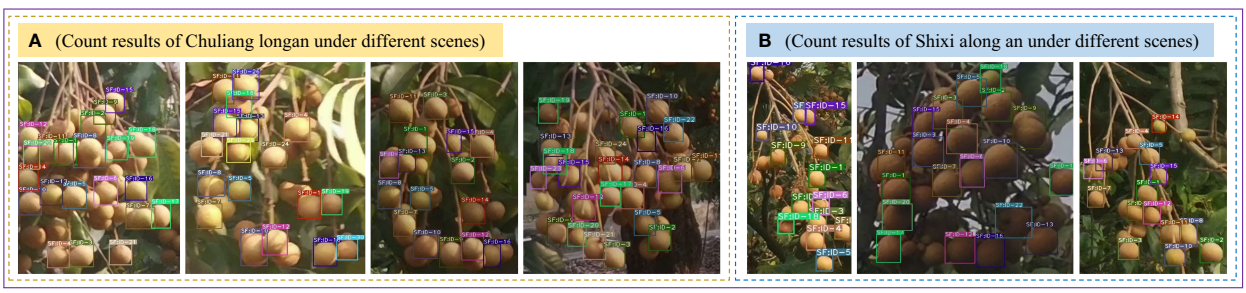


FIGURE 12
Single_Fruit counting results of the SF-YD model under different scenes.

10.3389/fpls.2023.1132909

TABLE 4 Sample number information of the two identification models and manual statistics.

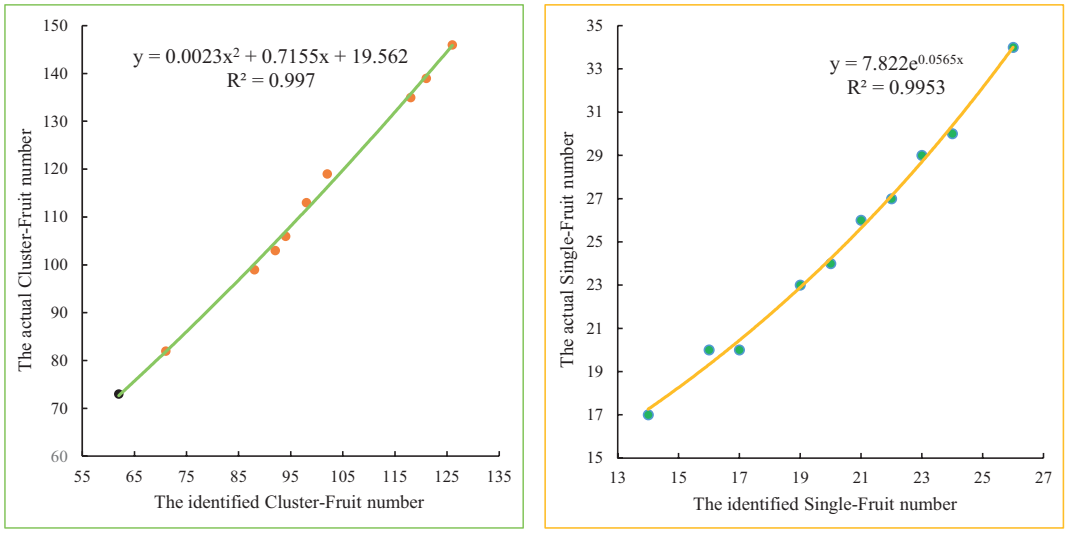
Class	Identified value (Cluster_Fruit)	Actual value (Cluster_Fruit)	Class	Identified value (Single_Fruit)	Actual value (Single_Fruit)
1	275	299	1	26	34
2	283	312	2	21	26
3	261	279	3	19	23
4	213	241	4	23	29
5	195	212	5	20	24
6	153	179	6	24	30
7	182	198	7	22	27
8	93	116	8	17	20
9	113	129	9	16	20
10	77	91	10	14	17

The real value of Cluster_Fruits and Single_Fruits are obtained by manual statistics, which is a multiangle and full-range process. However, the UAV collects video images of the fruit tree canopy from the front angle and can only obtain Cluster_Fruit and Single_Fruit images outside the tree canopy.

According to the statistics of horticulture experts, the average fruit weight of the Chuliang longan variety is 13 g and that of the Shixia longan variety is 8 g. After using the method proposed in this paper to obtain the number of Cluster_Fruits on a single fruit tree and the number of Single_Fruits on each Cluster_Fruit, the yield data of a single longan tree can be obtained by using the yield estimation strategy for a single fruit tree in Section 3.4.

Conclusion

In a complex longan orchard, fruit grow in clusters, and the shapes of Cluster_Fruits vary widely. It is difficult to estimate the yield of a single fruit tree simply by counting the number of Cluster_Fruits. Although the shapes Single_Fruits are relatively consistent, their shapes are small, so it is difficult to accurately count the number Single_Fruits directly by image analysis. Therefore, the yield estimation strategy based on UAV images proposed in this paper is of great significance and can improve the accuracy and efficiency of the yield statistics obtained for each fruit tree.

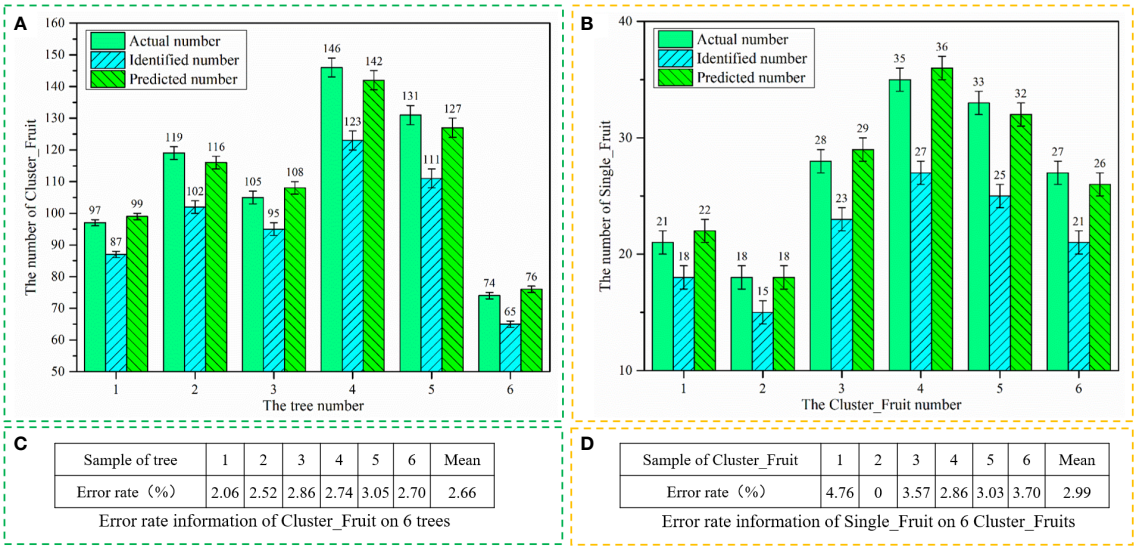


A (The fitting results of the actual and identified Cluster-Fruit number)

B (The fitting results of the actual and identified Single-Fruit number)

FIGURE 13
The fitting results of the actual and identified numbers of fruits.

10.3389/fpls.2023.1132909



C

Sample of tree	1	2	3	4	5	6	Mean
Error rate（%）	2.06	2.52	2.86	2.74	3.05	2.70	2.66

D

Sample of Cluster_Fruit	1	2	3	4	5	6	Mean
Error rate（%）	4.76	0	3.57	2.86	3.03	3.70	2.99

FIGURE 14
Statistical information of Cluster_Fruits and Single_Fruits. **(A)** The actual numbers, identified numbers and predicted numbers of Cluster_Fruits on six longan trees, **(B)** The actual numbers, identified numbers and predicted numbers of Single_Fruits on six Cluster_Fruits, **(C)** Error rate information of Cluster_Fruit on six trees, **(D)** Error rate information of Single_Fruit on six Cluster_Fruits.

In this study, a method based on UAV images and computer vision technology is proposed to estimate the yield of a single longan fruit tree. First, a UAV is used to collect video images of the fruit tree canopy, and after preprocessing the images, two datasets are constructed, and the targets of the datasets are manually marked. Then, the CF-YD and SF-YD models are constructed to identify Cluster_Fruits and Single_Fruits, respectively, which realizes the task of automatically identifying the number of targets directly from each image. Finally, to further predict the yield of a single longan fruit tree accurately, two models for estimating the numbers of Cluster_Fruits and Single_Fruits are proposed, and two fitting equations are established for determining the actual number and predicted number of Cluster_Fruits on a single fruit tree and the number of Single_Fruits on a single Cluster_Fruit, and the models are tested and verified in real orchards. This study can quickly and accurately estimate the yield of a single fruit tree, which can not only provide guidance for the production management and market pricing of longan orchards but also improve the efficiency of deploying harvesting robots and transportation robots, which is conducive to maximizing the economic benefits of orchards. The research in this paper can apply UAV image migration to the harvests of clustered fruits such as grapes and Cerasus pseudocerasus and promote the development of smart agriculture and unmanned farms.

Since most longan orchards are currently unstructured, this work still has some limitations, and the target detection and tracking abilities of the proposed method need to be further improved. In this study, the UAV mainly collects canopy images of longan fruit trees from the perspective of elevation but cannot obtain all-around images inside the canopies of fruit trees. Therefore, there is an error between the collected data and the real values. In future research, we will first consider the use of UAV to automatically plan flight routes in order to obtain orchard canopy images more easily. Secondly, an image analysis processor will be built on the UAV to calculate the output of fruit trees in real time. Finally, the result data of artificial statistics will continue to be added to further improve the accuracy of the fitting equation prediction quantity. In addition, the research objects will be expanded to more longan varieties in the future. In future work, we will continue to optimize the details of the solution to promote the development of smart agriculture.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

DL and XS conceived the study and wrote the paper. YJ, ZY, PL, YC, HZ, ZZ, and KW participated in the experiment and analyzed the experimental data. JL and LS supervised the manuscript and made valuable inputs. All authors contributed to the article and approved the submitted version.

Funding

This research is supported by the earmarked fund for the Laboratory of Lingnan Modern Agriculture Project (NZ2021009),

the open competition program of top ten critical priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan of Guangdong Province (2022SDZG03), the China Agriculture Research System (No. CARS-32-11), the Special Project of Rural Vitalization Strategy of Guangdong Academy of Agricultural Sciences (No. TS-1-4), and the Guangdong Provincial Modern Agricultural Industry Technology System (No. 2021KJ123).

Acknowledgments

We are very grateful to Guangdong Academy of Agricultural Sciences for providing us with the site for collecting experimental data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alpaydin, E. (2016). Neural networks and deep learning. *machine learning: The new AI*. MIT Press. Cambridge, Massachusetts, USA. https://ieeexplore.ieee.org/document/7845182

Anagnostis, A., Tagarakis, A. C., Asiminari, G., Papageorgiou, E., Kateris, D., Moshou, D., et al. (2021). A deep learning approach for anthracnose infected trees classification in walnut orchards. *Comput. Electron. Agric.* 182. doi: 10.1016/j.compag.2021.105998

Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B. (2016). Simple Online and Realtime Tracking. *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, pp. 3464–3468. doi: 10.1109/ICIP.2016.7533003

Bochkovskiy, A., Wang, C. Y., and Liao, H. (2020). YOLOv4: Optimal speed and accuracy of object detection. doi: 10.48550/arXiv.2004.10934

da Silva, C. B., Bianchini, V. D. M., de Medeiros, A. D., de Moraes, M. H. D., Marassi, A. G., and Tannus, A. (2021). A novel approach for jatropha curcas seed health analysis based on multispectral and resonance imaging techniques. *Ind. Crops Prod.* 161. doi: 10.1016/j.indcrop.2020.113186

de Medeiros, A. D., Bernardes, R. C., da Silva, L. J., de Freitas, B. A. L., Dias, D. C. F. D., and da Silva, C. B. (2021). Deep learning-based approach using X-ray images for classifying crambe abyssinica seed quality. *Ind. Crops Prod.* 164. doi: 10.1016/j.indcrop.2021.113378

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., and Sun, J. (2021). RepVGG: Making VGG-style ConvNets Great Again. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 13728–13737. doi: 10.1109/CVPR46437.2021.01352.

Feng, A. J., Zhou, J. F., Vories, E., and Sudduth, K. A. (2020a). Evaluation of cotton emergence using UAV-based imagery and deep learning. *Comput. Electron. Agric.* 177. doi: 10.1016/j.compag.2020.105711

Feng, A. J., Zhou, J. F., Vories, E. D., Sudduth, K. A., and Zhang, M. N. (2020b). Yield estimation in cotton using UAV-based multi-sensor imagery. *Biosyst. Eng.* 193, 101–114. doi: 10.1016/j.biosystemseng.2020.02.014

Flores, P., Zhang, Z., Igathinathane, C., Jithin, M., Naik, D., Stenger, J., et al. (2021). Distinguishing seedling volunteer corn from soybean through greenhouse color, color-infrared, and fused images using machine and deep learning. *Ind. Crops Prod.* 161. doi: 10.1016/j.indcrop.2020.113223

Gao, F. F., Fu, L. S., Zhang, X., Majeed, Y., Li, R., Karkee, M., et al. (2020). Multi-class fruit-on-plant detection for apple in SNAP system using faster r-CNN. *Comput. Electron. Agric.* 176. doi: 10.1016/j.compag.2020.105634

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. doi: 10.48550/arXiv.2107.08430

Girshick, R. (2015b). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440–1448. doi: 10.1109/iccv.2015.169

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2015a). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (1), 142–158. doi: 10.1109/tpami.2015.2437384

He, L., Fang, W., Zhao, G., Wu, Z., Fu, L., Li, R., et al. (2022). Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. *Comput. Electron. Agric.* 195. doi: 10.1016/j.compag.2022.106812

He, Z. L., Xiong, J. T., Chen, S. M., Li, Z. X., Chen, S. F., Zhong, Z., et al. (2020). A method of green citrus detection based on a deep bounding box regression forest. *Biosyst. Eng.* 193, 206–215. doi: 10.1016/j.biosystemseng.2020.03.001

He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (9), 1904–1916. doi: 10.1109/tpami.2015.2389824

Jaisin, C., Pathaveerat, S., and Terdwongworakul, A. (2013). Determining the size and location of longans in bunches by image processing technique. *Maejo Int. J. Sci. Technol.* 7 (3), 444–455. doi: 10.14456/mijst.2013.37

Jiang, T., and Cheng, J. (2019). Target Recognition Based on CNN with LeakyReLU and PReLU Activation Functions. *2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, Beijing, China, pp. 718–722. doi: 10.1109/SDPC.2019.00136.

Koirala, A., Walsh, K. B., Wang, Z. L., and McCarthy, C. (2019). Deep learning - method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* 162, 219–234. doi: 10.1016/j.compag.2019.04.017

Li, D. H., Sun, X. X., Elkhouchlaa, H., Jia, Y. H., Yao, Z. W., Lin, P. Y., et al. (2021). Fast detection and location of longan fruits using UAV images. *Comput. Electron. Agric.* 190. doi: 10.1016/j.compag.2021.106465

Li, D. H., Sun, X. X., Lv, S. P., Elkhouchlaa, H., Jia, Y. H., Yao, Z. W., et al. (2022). A novel approach for 3D localization of branch picking points based on deep learning applied to fruit picking UAVs. *Comput. Electron. Agric.* 199. doi: 10.1016/j.compag.2022.107191

Liang, C. X., Xiong, J. T., Zheng, Z. H., Zhong, Z., Li, Z. H., Chen, S. M., et al. (2020). A visual detection method for nighttime litchi fruits and fruiting stems. *Comput. Electron. Agric.* 169. doi: 10.1016/j.compag.2019.105192

Lin, T. Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936–944. doi: 10.1109/CVPR.2017.106

Lin, G., Tang, Y., Zou, X., Cheng, J., and Xiong, J. (2020). Fruit detection in natural environment using partial shape matching and probabilistic hough transform. *Precis. Agric.* 21 (1), 160–177. doi: 10.1007/s11119-019-09662-w

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). Path aggregation network for instance segmentation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 8759–8768. doi: 10.1109/CVPR.2018.00913

Ma, L., Liu, Y., Zhang, X. L., Ye, Y. X., Yin, G. F., and Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *Isprs J. Photogramm. Remote Sens.* 152, 166–177. doi: 10.1016/j.isprsjprs.2019.04.015

Marani, R., Milella, A., Petitti, A., and Reina, G. (2021). Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera. *Precis. Agric.* 22 (2), 387–413. doi: 10.1007/s11119-020-09736-0

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., et al. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. United States America* 115 (25), E5716–E5725. doi: 10.1073/pnas.1719367115

Paoletti, M. E., Haut, J. M., Plaza, J., and Plaza, A. (2019). Deep learning classifiers for hyperspectral imaging: A review. *Isprs J. Photogramm. Remote Sens.* 158, 279–317. doi: 10.1016/j.isprsjprs.2019.09.006

Pham, V. T., Herrero, M., and Hormaza, J. I. (2015). Phenological growth stages of longan (Dimocarpus longan) according to the BBCH scale. *Sci. Hortic.* 189, 201–207. doi: 10.1016/j.scienta.2015.03.036

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779–788. doi: 10.1109/CVPR.2016.91

Redmon, J., and Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv e-prints*. doi: 10.48550/arXiv.1804.02767

Ren, S. Q., He, K. M., Girshick, R., and Sun, J. (2017). Faster r-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6), 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Singh, P., Verma, A., and Alex, J. S. R. (2021). Disease and pest infection detection in coconut tree through deep learning techniques. *Comput. Electron. Agric.* 182. doi: 10.1016/j.compag.2021.105986

Sumesh, K. C., Ninsawat, S., and Som-ard, J. (2021). Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. *Comput. Electron. Agric.* 180. doi: 10.1016/j.compag.2020.105903

Tang, Y. C., Zhou, H., Wang, H. J., and Zhang, Y. Q. (2023). Fruit detection and positioning technology for a camellia oleifera c. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Syst. Appl.* 211, 118573. doi: 10.1016/j.eswa.2022.118573

Tetila, E. C., Machado, B. B., Menezes, G. K., Oliveira, A. D., Alvarez, M., Amorim, W. P., et al. (2020). Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* 17 (5), 903–907. doi: 10.1109/LGRS.2019.2932385

Vanegas, F., Bratanov, D., Powell, K., Weiss, J., and Gonzalez, F. (2018). A novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data. *Sensors* 18 (1). doi: 10.3390/s18010260

Wang, C. Y., Bochkovskiy, A., and Liao, H. (2021). "Scaled-YOLOv4: Scaling cross stage partial network," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Wang, C. Y., Bochkovskiy, A., and Liao, H. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv e-prints*. doi: 10.48550/arXiv.2207.02696

Wang, C. L., Luo, T. H., Zhao, L. J., Tang, Y. C., and Zou, X. J. (2019). Window zooming–based localization algorithm of fruit and vegetable for harvesting robot. *IEEE Access* 7, 103639–103649. doi: 10.1109/access.2019.2925812

Wojke, N., Bewley, A., and Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, pp. 3645–3649. doi: 10.1109/ICIP.2017.8296962

Wu, F. Y., Duan, J. L., Ai, P. Y., Chen, Z. Y., Yang, Z., and Zou, X. J. (2022). Rachis detection and three-dimensional localization of cut off point for vision-based banana robot. *Comput. Electron. Agric.* 198, 107079. doi: 10.1016/j.compag.2022.107079

Xiong, J. T., He, Z. L., Lin, R., Liu, Z., Bu, R. B., Yang, Z. G., et al. (2018). Visual positioning technology of picking robots for dynamic litchi clusters with disturbance. *Comput. Electron. Agric.* 151, 226–237. doi: 10.1016/j.compag.2018.06.007

Xiong, J. T., Liu, Z., Chen, S. M., Liu, B. L., Zheng, Z. H., Zhong, Z., et al. (2020). Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method. *Biosyst. Eng.* 194, 261–272. doi: 10.1016/j.biosystemseng.2020.04.006

Zhong, Z., Xiong, J. T., Zheng, Z. H., Liu, B. L., Liao, S. S., Huo, Z. W., et al. (2021). A method for litchi picking points calculation in natural environment based on main fruit bearing branch detection. *Comput. Electron. Agric.* 189. doi: 10.1016/j.compag.2021.106398

Zhou, Y. H., Tang, Y. C., Zou, X. J., Wu, M. L., Tang, W., Meng, F., et al. (2022). Adaptive active positioning of camellia oleifera fruit picking points: Classical image processing and YOLOv7 fusion algorithm. *Appl. Sciences-Basel* 12 (24), 12959. doi: 10.3390/app122412959

Zhou, J., Zhou, J. F., Ye, H., Ali, M. L., Nguyen, H. T., and Chen, P. Y. (2020). Classification of soybean leaf wilting due to drought stress using UAV-based imagery. *Comput. Electron. Agric.* 175. doi: 10.1016/j.compag.2020.105576

Zhuang, J. J., Hou, C. J., Tang, Y., He, Y., Guo, Q. W., Zhong, Z. Y., et al. (2019). Computer vision-based localisation of picking points for automatic litchi harvesting applications towards natural scenarios. *Biosyst. Eng.* 187, 1–20. doi: 10.1016/j.biosystemseng.2019.08.016

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 March 2023
DOI 10.3389/fpls.2023.1150855

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Chao Wang,
Chengdu University of Information Technology, China
Alireza Sanaeifar,
Zhejiang University, China

*CORRESPONDENCE
Haiguang Wang
wanghaiguang@cau.edu.cn

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 25 January 2023
ACCEPTED 03 March 2023
PUBLISHED 17 March 2023

CITATION
Jiang Q, Wang H and Wang H (2023) Severity assessment of wheat stripe rust based on machine learning.
Front. Plant Sci. 14:1150855.
doi: 10.3389/fpls.2023.1150855

COPYRIGHT
© 2023 Jiang, Wang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Severity assessment of wheat stripe rust based on machine learning

Qian Jiang, Hongli Wang and Haiguang Wang*

College of Plant Protection, China Agricultural University, Beijing, China

Introduction: The accurate severity assessment of wheat stripe rust is the basis for the pathogen-host interaction phenotyping, disease prediction, and disease control measure making.

Methods: To realize the rapid and accurate severity assessment of the disease, the severity assessment methods of the disease were investigated based on machine learning in this study. Based on the actual percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves of each severity class of the disease, obtained after the image segmentation operations on the acquired single diseased wheat leaf images and the pixel statistics operations on the segmented images by using image processing software, under two conditions of considering healthy single wheat leaves or not, the training and testing sets were constructed by using two modeling ratios of 4:1 and 3:2, respectively. Then, based on the training sets, two unsupervised learning methods including *K*-means clustering algorithm and spectral clustering and three supervised learning methods including support vector machine, random forest, and *K*-nearest neighbor were used to build severity assessment models of the disease, respectively.

Results: Regardless of whether the healthy wheat leaves were considered or not, when the modeling ratios were 4:1 and 3:2, satisfactory assessment performances on the training and testing sets can be achieved by using the optimal models based on unsupervised learning and those based on supervised learning. In particular, the assessment performances obtained by using the optimal random forest models were the best, with the accuracies, precisions, recalls, and F1 scores for all the severity classes of the training and testing sets equal to 100.00% and the overall accuracies of the training and testing sets equal to 100.00%.

Discussion: The simple, rapid, and easy-to-operate severity assessment methods based on machine learning were provided for wheat stripe rust in this study. This study provides a basis for the automatic severity assessment of wheat stripe rust based on image processing technology, and provides a reference for the severity assessments of other plant diseases.

KEYWORDS

wheat stripe rust, severity, disease assessment, image processing, machine learning, unsupervised learning, supervised learning

1 Introduction

Wheat stripe rust (wheat yellow rust), caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*), is a devastating epidemic disease, which is widely distributed in wheat growing areas worldwide (Li and Zeng, 2002; Line, 2002; Chen, 2005; Wellings, 2011; Wang et al., 2014; Ali et al., 2017; Figueroa et al., 2018). As an air-borne fungal disease, wheat stripe rust can cause multiple uredinia on wheat leaves and a large number of *Pst* urediospores can be produced and released. The dispersal of *Pst* urediospores by wind can cause large-scale epidemics of the disease in wheat growing areas, which can seriously affect the yield and quality of wheat (Li and Zeng, 2002; Chen, 2005; Chen et al., 2014; Wang et al., 2014). In China, wheat stripe rust, one of the most important wheat diseases, has been a serious threat to wheat production safety because of its high epidemic frequency and severe destructiveness, and the pandemics of the disease have occurred many times, resulting in huge wheat yield losses (Li and Zeng, 2002; Chen et al., 2014; Wang et al., 2014; Liu et al., 2022). There are three epidemiological region systems of wheat stripe rust in China, including the northern China–north-western China–the middle and lower reaches of the Yangtze River epidemiological region system, Xinjiang epidemiological region system, and Yunnan epidemiological region system, and the disease cycles are completed by air dispersals of *Pst* urediospores in each epidemiological region system or across the epidemiological region systems (Li and Zeng, 2002; Chen et al., 2014; Wang et al., 2014). To effectively control the occurrence and epidemics of wheat stripe rust and to achieve sustainable management of the disease, it is of great significance to carry out surveys and monitoring of the disease.

Severity is a key indicator to describe the disease intensity of an investigated plant unit (a plant or a plant part such as leaf, fruit, or stem), and different severity levels of a plant disease are usually classified based on the ratios of the diseased areas of the investigated plant units to the areas of the corresponding whole investigated plant units (Nutter et al., 1991; Bock et al., 2022b). Disease severity assessment should be performed according to the severity grading standard of a plant disease to ensure the standardization and integrity of the obtained data. Severity is an important indicator to be determined in surveys and monitoring of wheat stripe rust. In China, the severity assessment of wheat stripe rust should be conducted according to the Rules for Monitoring and Forecast of the Wheat Stripe Rust (*Puccinia striiformis* West.) (National Standard of the People's Republic China, GB/T 15795–2011). In this standard, a total of eight severity classes including 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, are classified based on the percentages of the lesion areas (Generally, the lesion area refers to the area covered by all the uredinia on a diseased wheat leaf.) in the areas of the corresponding whole single diseased wheat leaves. According to this severity grading standard, the disease intensity of a single diseased wheat leaf is treated as its nearest severity class based on the nearest-neighbor rule, and as the disease intensity is lower than the severity class of 1%, it is classified as the severity class of 1%.

At the present time, the severity assessments of wheat stripe rust are implemented mainly by using the visual observation method. This kind of artificial method requires assessors/raters with rich experience, and it is time-consuming and laborious. Most importantly, it is difficult to accurately estimate the percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves according to the severity grading standard of wheat stripe rust (Jiang et al., 2022), thus it is difficult to obtain accurate severity assessment results by using this method. Because the investigation of disease incidence only needs to determine whether the wheat leaves are diseased or not and it is easier to accurately investigate or assess disease incidence than disease severity, it was reported that disease incidence could be used to estimate the severity of wheat stripe rust (Dong et al., 1990). However, the quantitative relationship between incidence and severity (*I*-*S* relationship) is affected by the values of disease incidences, the distribution of lesions on wheat plants, wheat resistance to *Pst*, and so on (Dong et al., 1990), leading to great limitations to the practical applications of the method by utilizing incidence to estimate severity.

The assessment of plant disease severity based on information technology has been paid more and more attention (Bock et al., 2010; Li et al., 2015; Barbedo, 2016; Mahlein, 2016; Bock et al., 2022a). The rapid development of information technology has promoted the applications of image processing technology (Bao et al., 2021; Jiang et al., 2021; Jiang et al., 2022), remote sensing technology (Huang et al., 2004; Wang et al., 2007; Zhao et al., 2014; Wang et al., 2016), and near infrared spectroscopy technology (Li et al., 2015) in severity assessment of wheat stripe rust. Nevertheless, the methods for severity assessment of wheat stripe rust based on remote sensing technology and near infrared spectroscopy technology are rarely used in practical disease surveys and monitoring due to the high prices of the required instruments and the need to further improve the applicability of the related methods in practical wheat production. Studies on severity assessment of wheat stripe rust based on image processing technology are increasing (Bao et al., 2021; Jiang et al., 2021; Jiang et al., 2022).

At present, there are two main methods based on image processing technology to be utilized to carry out severity assessment of wheat stripe rust. One method is to directly build the severity assessment models of wheat stripe rust based on the extracted features (e.g., color, shape, and texture features) from disease images and then to carry out disease severity assessment by using the built models (Bao et al., 2021). The other method is to use image processing technology to obtain the actual percentages of the lesion areas in the areas of the whole single diseased wheat leaves, then to directly compare the actual percentages to the percentages for the eight severity classes in the severity grading standard of wheat stripe rust, and to obtain the severity classes of the corresponding diseased wheat leaves finally (Jiang et al., 2021). However, the percentage of the lesion area in the area of a whole single diseased wheat leaf corresponding to a severity class in the severity grading standard of wheat stripe rust is not the actual percentage of the lesion area in the area of the whole single diseased wheat leaf, and importantly, there is a great difference between them. This has been verified by the studies conducted by Shang et al. (1990) and Jiang et al. (2022). By using a method based on the uredinium parameters in combination with actually measuring the

amplified image of the selected wheat leaf with the most severe disease symptom in the field, Shang et al. (1990) determined actual coverage rates of the *Pst* uredinia corresponding to the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, which were 0.35%, 1.75%, 3.5%, 7%, 14%, 21%, 28%, and 35%, respectively. By using image processing software to perform the operations of image segmentation and pixel statistics based on the acquired single diseased wheat leaf images, Jiang et al. (2022) obtained the ranges of the actual percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves for the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, which were [0.06%, 0.78%], [0.85%, 1.64%], [1.73%, 3.29%], [3.65%, 6.31%], [6.76%, 13.88%], [14.22%, 18.43%], [18.90%, 24.15%], and [24.54%, 36.49%], respectively. The severity assessment method by directly comparing the actual percentage of lesion area in the area of a whole single diseased wheat leaf to the percentages of lesion areas of the eight severity classes in the severity grading standard of wheat stripe rust, can lead to great errors in severity assessments, which will greatly influence the accurate severity assessments of the disease. Therefore, it is difficult to accurately assess the severity of wheat stripe rust by directly comparing the actual percentage of lesion area in the area of a whole single diseased wheat leaf to the percentages of lesion areas of the eight severity classes in the severity grading standard of the disease. To accurately carry out severity assessment of wheat stripe rust based on the actual percentages of lesion areas in the areas of the corresponding whole single diseased wheat leaves, Jiang et al. (2022) proposed two reference-range-based methods for severity assessment of wheat stripe rust, and satisfactory results with the assessment accuracies not lower than 85% were achieved by using the determined reference ranges to conduct severity assessments of the disease. However, the methods for severity assessment of wheat stripe rust proposed by Jiang et al. (2022) need to compare the actual percentages of lesion areas in the areas of the whole single diseased wheat leaves to the upper and lower limits of the determined reference range of the actual percentages of lesion areas for each severity class, and then the severity classes of the single diseased wheat leaves to be assessed can be determined accordingly.

To timely and accurately obtain the severity information of wheat stripe rust, it is necessary to develop a simple, rapid, accurate, and easy-to-operate severity assessment method for wheat stripe rust. On the basis of the study conducted by Jiang et al. (2022), the severity assessment methods based on machine learning were developed for wheat stripe rust in this study. The obtained actual percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves were clustered into different severity classes by using two unsupervised learning methods including *K*-means clustering algorithm and spectral clustering, respectively, and the built clustering models were treated as the severity assessment models of wheat stripe rust based on unsupervised learning. Simultaneously, the severity assessment models of wheat stripe rust were built with the obtained actual percentage data by using three supervised learning methods including support vector machine (SVM), random forest (RF), and *K*-nearest neighbor (KNN), respectively. To ensure that healthy wheat leaves could be assessed, severity assessment models of wheat stripe rust were also built by using the above five modeling methods under the condition of considering single healthy wheat leaves. Finally, all the built models were used to carry out severity assessments of the single wheat leaves. The goal of this study is to overcome the difficulties in severity assessment of wheat stripe rust and to solve the problem of low assessment accuracy caused by directly comparing the actual percentages of lesion areas in the areas of the single diseased wheat leaves to the percentages of lesion areas of the eight severity classes in the severity grading standard of the disease. This study will provide rapid and accurate severity assessment methods for wheat stripe rust based on the actual percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves, and also provide a reference and basis for the severity assessments of other plant diseases and the automatic assessments of plant disease severity.

2 Materials and methods

The main steps for developing the severity assessment methods of wheat stripe rust based on machine learning in this study are shown in Figure 1. Especially, to enable the built severity assessment models to be used for severity assessments of single wheat leaves including healthy leaves and to implement automatic disease assessments for acquired single wheat leaves in the future, in this study, the severity assessment of wheat stripe rust based on mechanical learning was investigated under two conditions of considering healthy single wheat leaves or not.

2.1 Disease images and image processing methods

The images of wheat stripe rust and the corresponding actual percentages of lesion areas in the areas of the single diseased wheat leaves used in this study were as same as those acquired by Jiang et al. (2022). Therefore, the images, data, and relevant methods used to obtain the images and data, are only briefly described here.

A total of 400 images of wheat stripe rust were used in this study, which were acquired by using 400 single diseased wheat leaves of eight severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100% collected on the wheat plants in Shangzhuang Experimental Station of China Agricultural University, Beijing, China and an artificial climate chamber in the Laboratory of Macro-Phytopathology, China Agricultural University, Beijing, China, according to the Rules for Monitoring and Forecast of the Wheat Stripe Rust (*Puccinia striiformis* West.) as described above. For each severity class, 50 images were acquired (one image per single diseased wheat leaf) by using a Nikon D700 digital camera (Nikon Corp., Tokyo, Japan), a HUAWEI P30 smartphone, or an iPhone 6S smartphone. The sizes of the images acquired with the Nikon D700 digital camera, the HUAWEI P30 smartphone, and the iPhone 6S smartphone were 4256×2832, 3648×2736, and 4032×3024 pixels, respectively, and all the disease images were in the JEPG format. In the Adobe Photoshop 2022 software (Adobe Systems Incorporated, San Jose, CA, USA), leaf region segmentation and lesion region segmentation of the single diseased wheat leaf images were performed, and then the pixel number of the whole leaf region and the pixel number of the lesion

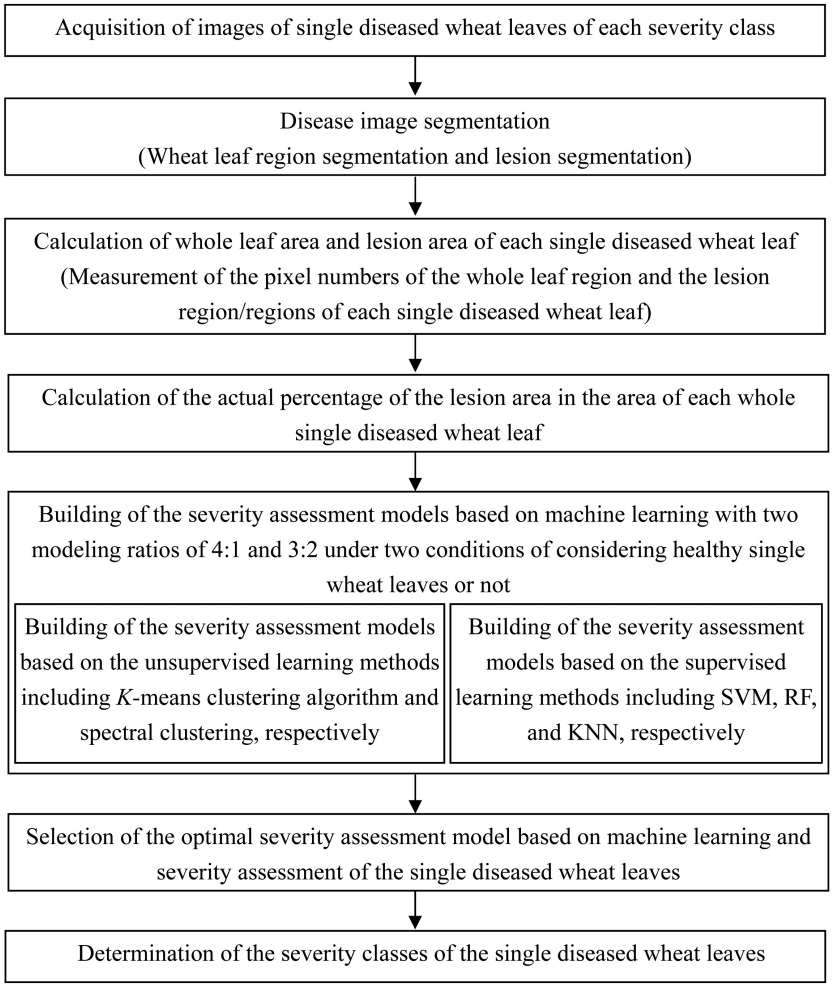


FIGURE 1
Work flow diagram for investigation of severity assessment methods of wheat stripe rust based on machine learning.

region/regions for each single diseased wheat leaf were obtained by pixel statistics. Based on the pixel numbers of the whole leaf region and the lesion region/regions, the actual percentage of the lesion area in the area of the whole single diseased wheat leaf was calculated.

2.2 Building of the severity assessment models of wheat stripe rust based on machine learning under the condition without considering single healthy wheat leaves

The eight severity classes of wheat stripe rust were regarded as eight categories, i.e., the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100% were set to the severity categories of 1, 2, 3, 4, 5, 6, 7, and 8, respectively. Each category was composed of 50 single diseased wheat leaf specimens of wheat stripe rust. For each category, the actual percentages of the lesion areas in the corresponding whole leaf areas for the 50 specimens were sorted from large to small, and then, the specimens were sampled to construct the training set and the testing set by using the system sampling method with the modeling ratio of the number of specimens in the training set to the number of specimens in the testing set equal to 4:1 or 3:2. Subsequently, for each modeling ratio (4:1 or 3:2), the training sets of the eight categories (severity classes) were merged into new training set (Train840 or Train830) and the testing sets of the eight categories were merged into testing set (Test810 or Test820). For the modeling ratio of 4:1, the training set Train840 consisted of 320 specimens, and the corresponding testing set Test810 consisted of 80 specimens. For the modeling ratio of 3:2, the training set Train830 and the corresponding testing set Test820 consisted of 240 specimens and 160 specimens, respectively.

2.2.1 Building of the severity assessment models of wheat stripe rust based on unsupervised learning under the condition without considering single healthy wheat leaves

The actual percentage of the lesion area in the whole wheat leaf area and the corresponding category number of each specimen in the training set Train840, the testing set Test810, the training set Train830, and the testing set Test820 were input into the data tables

in Microsoft Excel 2016, and the corresponding data tables were named train840, test810, train830, and test820, respectively. In each data table, the data in the first column were the actual percentages of the lesion areas in the corresponding whole wheat leaf areas and were recorded as X; the data in the second column were the corresponding severity category numbers and were recorded as Y.

Under the condition without considering single healthy wheat leaves, based on the actual percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves, the severity assessment models of wheat stripe rust were built with the two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering by using the programming language Python (version: 3.8.12) in the software Pycharm 2021.2.1, respectively.

The *K*-means clustering algorithm is an unsupervised learning clustering algorithm based on distance clustering. It separates the data samples into different categories by initializing centroids and constantly updating the clustering centroids, and the clustering will be accomplished until the data samples in each category does not change, the maximum number of iterations is reached, or the error is lower than the expected value (MacQueen, 1967). To build the severity assessment models of wheat stripe rust with the *K*-means clustering algorithm, firstly, the data X and Y were read from the data tables train840, test810, train830, and test820, respectively, by calling the read_excel() method from the Pandas library. Then, the reshape (-1,1) method was used to convert the data X from a one-dimensional array into a two-dimensional array with one column to meet the requirements of scikit-learn library (Pedregosa et al., 2011) for input data. Subsequently, the severity assessment models of wheat stripe rust were built with different parameters by calling the module sklearn.cluster from the scikit-learn library. The method fit() was used to train the models based on the data X of the training sets. The labels over the data X of a training set after clustering were viewed by using the labels_ attribute and were recorded as train_label, and the number of each label was counted by calling the value_counts() method from the Pandas library. The method predict() was used to predict the labels of the data X of the corresponding testing sets by using the trained models. The predicted labels over the data X of a testing set were recorded as test_label, and the number of each label was counted by calling the value_counts() method from the Pandas library. When the severity assessment model of wheat stripe rust was built with the *K*-means clustering algorithm based on the training set Train840, the parameter n_clusters, that is, the number of clusters to form (i.e., the number of centroids to generate), was set to 8; the parameter random_state, that is, the generator used to determine random number generation for centroid initialization, was set to 10; the parameter n_init, that is, the number of times for random initialization, was set to 1; the parameter init, that is, the method for initialization, was set to 'random', which means that the initial centroids were randomly selected from the training data; and the default values were used for the other parameters. When the severity assessment model of wheat stripe rust was built with the *K*-means clustering algorithm based on the training set Train830, the parameter n_clusters was set to 8, the parameter random_state was set to 8, the parameter n_init was set to 1, the parameter init was set to 'random', and the default values were used for the other parameters.

Spectral clustering is an unsupervised learning clustering algorithm based on graph theory, by firstly establishing the Laplacian matrix of the data of samples, then calculating the eigenvalues and eigenvectors of the matrix, subsequently constructing the eigenvector space, and finally, clustering the eigenvectors to accomplish the clustering of the data samples with a traditional clustering algorithm such as the *K*-means algorithm (Shi and Malik, 2000; von Luxburg, 2007). To build the severity assessment models of wheat stripe rust with spectral clustering, the data X and Y were firstly read from the data tables train840, test810, train830, and test820, respectively, by calling the method read_excel () from the Pandas library, and then, the reshape(-1,1) method was used to convert the data X from a one-dimensional array into a two-dimensional array with one column to meet the requirements of scikit-learn library (Pedregosa et al., 2011) for input data. Subsequently, by calling the module sklearn.cluster from the scikit-learn library, the severity assessment models of wheat stripe rust were built with spectral clustering with different parameters. The method fit() was used to train the models based on the data X of the training sets. The labels over the data X of a training set after clustering were found in the labels_ attribute and were recorded as train_label, and the number of each label was counted by using the method value_counts() of the Pandas library. The method fit_predict() was used to predict the labels of the data X of the corresponding testing sets by using the trained models. The predicted labels over the data X of a testing set were recorded as test_label, and the counts of the unique labels were obtained by using the method value_counts() of the Pandas library. To build the severity assessment model of wheat stripe rust with spectral clustering based on the training set Train840, the parameter n_clusters, that is, the number of clustering dimensions (the number of clusters to form), was set to 8; the parameter affinity, that is, the method used to construct the affinity matrix, was set to 'nearest_neighbors', which means that the affinity matrix was constructed by using the nearest neighbors method; the parameter n_neighbors, that is, the number of neighbors used to construct the affinity matrix by using the nearest neighbors method, was set to 10; the parameter assign_labels, that is, the strategy used to assign labels in the embedding space, was set to 'discretize', which means that discretization approach was used to assign labels in the embedding space; and the default values were used for the other parameters. To build the severity assessment model of wheat stripe rust with spectral clustering based on the training set Train830, the parameter n_clusters was set to 8, the parameter affinity was set to 'nearest_neighbors', the parameter n_neighbors was set to 7, the parameter assign_labels was set to 'discretize', and the default values were used for the other parameters.

2.2.2 Building of the severity assessment models of wheat stripe rust based on supervised learning under the condition without considering single healthy wheat leaves

Under the condition without considering single healthy wheat leaves, based on the actual percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves, the severity

assessment models of wheat stripe rust were built with the three supervised learning methods including SVM (Cortes and Vapnik, 1995), RF (Breiman, 2001), and KNN (Cover and Hart, 1967), respectively.

By using the C-SVM in the LIBSVM-3.23 package developed by Chih-Jen Lin Group from National Taiwan University, Taiwan, China (Chang and Lin, 2011), the severity assessment SVM models of wheat stripe rust were built in this study. When building a SVM model for severity assessment of wheat stripe rust, the RBF kernel function was used. The grid search algorithm was utilized to determine the optimal values for both penalty parameter C and kernel function parameter g by searching in a range of 2^{-10} – 2^{10} with the searching step equal to 0.4. By using the 3-fold cross-validation method, when the assessment accuracy was the highest for the training set, the corresponding values of the parameters C and g were regarded as the optimal parameters to build the optimal SVM model for severity assessment of wheat stripe rust.

Building of the severity assessment RF models of wheat stripe rust was implemented in the software MATLAB R2019b (MathWorks, Natick, MA, USA). When building the severity assessment RF models of wheat stripe rust based on a training set, the number of decision trees, a key parameter for modeling, was set to 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100, respectively, and the other parameters with the default values were used. The model performance for severity assessment of wheat stripe rust was used to determine the optimal number of decision trees for modeling. Then, the determined optimal number of decision trees and the other parameters with the default values were used to build the optimal RF model for severity assessment of wheat stripe rust.

Building of the severity assessment KNN models of wheat stripe rust was implemented by using the KNN classifier in the software MATLAB R2019b. When building the severity assessment KNN models of wheat stripe rust based on a training set, Euclidean distance was selected as the default distance metric measure, the values of the key parameter K were selected in a range of 1–20 with the searching step of 2, and the other parameters with the default values were used. According to the assessment accuracies of the models, the optimal value of K was determined when the assessment accuracy was the highest for the training set. Then, the optimal KNN model for severity assessment of wheat stripe rust was built with the optimal K and the other parameters with the default values.

2.3 Building of the severity assessment models of wheat stripe rust based on machine learning under the condition of considering single healthy wheat leaves

To ensure that the built severity assessment models of wheat stripe rust could be used to assess the specimens of healthy wheat leaves, the actual percentage of the lesion area to the whole leaf area of a single healthy wheat leaf was set to 0% (i.e., the corresponding severity class was set to 0%.), and 50 '0%' for healthy wheat leaves were added and then were numbered, respectively. The severity category of each healthy wheat leaf was set to 0, and the severity categories of the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100% were set to 1, 2, 3, 4, 5, 6, 7, and 8, respectively. Thus, the severity classes of wheat stripe rust were regarded as nine categories (0–8) and each category consisted of 50 single wheat leaf specimens. By using the method as described above, the actual percentages of the lesion areas in the corresponding whole leaf areas for the 50 specimens of each category were sorted from large to small, and then the specimens were sampled to construct the training set and the testing set by using the system sampling method with the modeling ratio of the number of specimens in the training set to the number of specimens in the testing set equal to 4:1 or 3:2. Subsequently, the training sets of the nine categories were merged into new training set Train940 or Train930, and the testing sets of the nine categories were merged into the corresponding new testing set Test910 or Test920. For the modeling ratio of 4:1, the training set Train940 and the corresponding testing set Test910 consisted of 360 specimens and 90 specimens, respectively. For the modeling ratio of 3:2, the training set Train930 and the corresponding testing set Test920 consisted of 270 specimens and 180 specimens, respectively.

2.3.1 Building of the severity assessment models of wheat stripe rust based on unsupervised learning under the condition of considering single healthy wheat leaves

By using the same method as described above, the data tables train940, test910, train930, and test920 were created based on the training set Train940, the testing set Test910, the training set Train930, and the testing set Test920, respectively. In each data table, the data in the first column recorded as X were the actual percentages of the lesion areas in the corresponding whole wheat leaf areas and that in the second column recorded as Y were the corresponding category numbers of disease severity.

By using the similar methods to build the severity assessment models of wheat stripe rust with the two unsupervised learning methods under the condition without considering single healthy wheat leaves as described above, the severity assessment models of wheat stripe rust under the condition of considering single healthy wheat leaves were built with the two unsupervised learning methods including the K-means clustering algorithm and spectral clustering, respectively, based on the actual percentages of the lesion areas in the areas of the corresponding whole single wheat leaves by using the programming language Python (version: 3.8.12) in the software Pycharm 2021.2.1. To build the severity assessment model of wheat stripe rust with the K-means clustering algorithm based on the training set Train940, the parameter n_clusters was set to 9, the parameter random_state was set to 8, the parameter n_init was set to 1, the parameter init was set to 'random', and the default values were used for the other parameters. To build the severity assessment model of wheat stripe rust with the K-means clustering algorithm based on the training set Train930, the parameter n_clusters was set to 9, the parameter random_state was set to 7, the parameter n_init was set to 1, the parameter init was set to 'random', and the default values were used for the other parameters. To build the severity assessment model of wheat stripe rust with spectral clustering based on the training set Train940, the parameter n_clusters was set to 9, the parameter affinity

was set to 'nearest_neighbors', the parameter n_neighbors was set to 10, the parameter assign_labels was set to 'discretize', and the default values were used for the other parameters. To build the severity assessment model of wheat stripe rust with spectral clustering based on the training set Train930, the parameter n_clusters was set to 9, the parameter affinity was set to 'nearest_neighbors', the parameter n_neighbors was set to 8, the parameter assign_labels was set to 'discretize', and the default values were used for the other parameters.

2.3.2 Building of the severity assessment models of wheat stripe rust based on supervised learning under the condition of considering single healthy wheat leaves

By using the same methods to build the severity assessment models of wheat stripe rust with the three supervised learning methods under the condition without considering single healthy wheat leaves as described above, the severity assessment models of wheat stripe rust under the condition of considering single healthy wheat leaves were built based on the actual percentages of the lesion areas in the areas of the corresponding whole single wheat leaves with the three supervised learning methods including SVM, RF, and KNN, respectively.

2.4 Performance evaluation of the built severity assessment models of wheat stripe rust

By using the severity assessment models of wheat stripe rust built with the two unsupervised learning methods (including the *K*-means clustering algorithm and spectral clustering) and the three supervised learning methods (including the SVM, RF, and KNN), the severity assessments of the corresponding training sets and testing sets were carried out. The accuracy, precision, recall, and F1 score of the severity assessments for the specimens of each severity class in the corresponding training sets and testing sets and the overall accuracies of the corresponding training sets and testing sets were calculated, respectively, aiming to evaluate the assessment performances of the built models and to choose the optimal severity assessment model of wheat stripe rust for each modeling ratio under the condition without considering single healthy wheat leaves or under the condition of considering single healthy wheat leaves. The accuracy for a severity class of wheat stripe rust is the percentage of the number of the correctly assessed single leaf specimens in the total number of single leaf specimens to be assessed. The precision is the percentage of the number of actual single leaf specimens at a severity class of wheat stripe rust in the number of the single leaf specimens assessed as the severity class. Recall is the percentage of the number of the single leaf specimens assessed as a severity class of wheat stripe rust in the number of the single leaf specimens actually at the severity class. F1 score is the harmonic mean of precision and recall. The overall accuracy is the percentage of the number of the single leaf specimens that are correctly assessed as the corresponding severity classes in the total number of single leaf specimens to be assessed. All the five evaluation indicators of severity assessment were calculated according to Formulas (1)–(5), respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\,\% \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\,\% \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\,\% \tag{3}$$

$$\text{F1 score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{|N|}\text{TP}_i}{\sum_{i=1}^{|N|}(\text{TP}_i + \text{FN}_i)} \times 100\,\% \tag{5}$$

where TP (true positive) is the number of the single leaf specimens actually at a severity class assessed as the severity class; FN (false negative) is the number of the single leaf specimens actually at the severity class assessed as the other severity classes; FP (false positive) is the number of the single leaf specimens actually at other severity classes assessed as the severity class; TN (true positive) is the number of the single leaf specimens actually at other severity classes correctly assessed as the corresponding other severity classes; N is the number of severity classes; and i is the ith severity class.

3 Results

3.1 Statistical results of the actual percentage data of the lesion areas in the areas of the corresponding whole single diseased wheat leaves at each severity class of wheat stripe rust

Statistical analysis of the data of the actual percentages of the lesion areas in the areas of the 50 corresponding whole single diseased wheat leaves at each severity class of wheat stripe rust, was conducted by using the UNIVARIATE procedure in the software SAS 9.4 (SAS Institute Inc. Cary, NC, USA). The minimum, maximum, mean, and standard deviation of the actual percentages of the lesion areas corresponding to each severity class of wheat stripe rust were obtained, as shown in Table 1. For the actual percentages of the lesion areas corresponding to the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, the minimum values were 0.06%, 0.85%, 1.73%, 3.65%, 6.76%, 14.22%, 18.90%, and 24.54%, respectively; the maximum values were 0.78%, 1.64%, 3.29%, 6.31%, 13.88%, 18.43%, 24.15%, and 36.49%, respectively; and the means were 0.40%, 1.27%, 2.50%, 4.92%, 9.88%, 16.61%, 21.24%, and 30.53%, respectively. The results showed that the actual percentages of the lesion areas corresponding to each severity class of wheat stripe rust were obviously lower than the percentage of the lesion area for the corresponding severity class in the severity grading standard of wheat stripe rust as described above.

TABLE 1 Statistics of the actual percentage data of the lesion areas in the areas of the whole diseased wheat leaves for the 50 acquired specimens corresponding to each severity class of wheat stripe rust including the minimum, maximum, mean, and standard deviation.

Severity class	Minimum of the actual percentages of the lesion areas	Maximum of the actual percentages of the lesion areas	Mean of the actual percentages of the lesion areas	Standard deviation of the actual percentages of the lesion areas
1%	0.06%	0.78%	0.40%	0.18%
5%	0.85%	1.64%	1.27%	0.23%
10%	1.73%	3.29%	2.50%	0.42%
20%	3.65%	6.31%	4.92%	0.77%
40%	6.76%	13.88%	9.88%	1.95%
60%	14.22%	18.43%	16.61%	1.20%
80%	18.90%	24.15%	21.24%	1.41%
100%	24.54%	36.49%	30.53%	3.17%

3.2 Severity assessment results obtained by using the severity assessment models of wheat stripe rust built based on the two unsupervised learning methods including the *k*-means clustering algorithm and spectral clustering under the condition without considering single healthy wheat leaves

Under the condition without considering single healthy wheat leaves, by using the severity assessment models of wheat stripe rust built based on the two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering, the severity assessments of the specimens of all the severity classes in the training and testing sets were carried out. The severity assessment results obtained by using the two unsupervised learning methods are shown in Tables 2, 3, respectively.

Under the condition without considering single healthy wheat leaves, when the modeling ratio was 4:1, the severity assessment model of wheat stripe rust was built by using the *K*-means clustering algorithm based on the training set Train840. By using the built model to conduct the severity assessments of the specimens in the training set Train840, the obtained results, as shown in Table 2, demonstrated that, for all the severity classes of wheat stripe rust, the lowest accuracy of 93.13% and the highest accuracy of 99.38% were obtained; the lowest and highest precisions of 71.43% and 100.00% were obtained, respectively; the lowest recall was 70.00%, and the highest recall was 100.00%; and the lowest F1 score was 73.17%, and the highest F1 score was 97.56%. By using the built model to conduct the severity assessments of the specimens in the testing set Test810, the results showed that, for all the severity classes of wheat stripe rust, the lowest and highest accuracies were 92.50% and 100.00%, respectively; the lowest values for precision, recall, and F1 score were all 70.00%, and the highest values for precision, recall, and F1 score were all 100.00%. There were large differences between the severity assessment results for each severity class in the training set Train840 or the testing set Test810 obtained by using the model built based on the *K*-means clustering algorithm. In detail, for all the severity classes, the obtained accuracies were very high; however, there were large differences in precision, recall, and F1 score. For the built severity assessment model, the overall accuracy of the training set Train840 was 87.81%, and that of the corresponding testing set Test810 was 87.50%.

Under the condition without considering single healthy wheat leaves, when the modeling ratio was 4:1, the severity assessment model of wheat stripe rust was built based on the training set Train840 by using spectral clustering. As shown in Table 3, the results obtained by using the built model to conduct the severity assessments of the specimens in the training set Train840 demonstrated that, for all the severity classes of wheat stripe rust, the lowest accuracy was 98.44% and the highest accuracy was 100.00%; except that the precisions for the severity classes of 5% and 80% were 88.89% and 93.02%, respectively, the precisions for all other severity classes were 100.00%; except that the recalls for the severity classes of 10% and 100% were 87.50% and 92.50%, respectively, the recalls for all other severity classes were 100.00%; and the lowest F1 score was 93.33%, and the highest F1 score was 100.00%. By using the built model to conduct the severity assessments of the specimens in the testing set Test810, the results showed that, except that the accuracies for the severity classes of 5% and 10% were both 98.75%, the accuracies for all other severity classes were 100.00%; except that the precision for the severity class of 5% was 90.91%, the precisions for all other severity classes were 100.00%; except that the recall for the severity class of 10% was 90.00%, the recalls for all other severity classes were 100.00%; except that the F1 scores for the severity classes of 5% and 10% were 95.24% and 94.74%, respectively, the F1 scores for all other severity classes were 100.00%. For the severity assessment model of wheat stripe rust built based on the training set Train840 by using spectral clustering, the overall accuracies of the training set used for modeling and the corresponding testing set Test810 were 97.50% and 98.75%, respectively. The results indicated that very good severity assessment results of the training and testing sets were achieved by using the severity assessment model built based on spectral clustering, when the modeling ratio was 4:1 under the condition without considering single healthy wheat leaves. There were relatively small differences among the severity assessment performances of the built model on all the severity classes of

10.3389/fpls.2023.1150855

TABLE 2 Severity assessment results of the single diseased wheat leaves with the actual percentages of lesion areas of all the severity classes of wheat stripe rust contained in the training and testing sets, obtained by using the severity assessment models of wheat stripe rust built based on the *K*-means clustering algorithm under the condition without considering single healthy wheat leaves.

Dataset	Severity category	Severity class	Accuracy	Precision	Recall	F1 score	Overall accuracy
Train840	1	1%	99.38%	95.24%	100.00%	97.56%	87.81%
	2	5%	96.25%	79.17%	95.00%	86.36%	
	3	10%	93.13%	71.43%	75.00%	73.17%	
	4	20%	93.75%	77.78%	70.00%	73.68%	
	5	40%	96.88%	100.00%	75.00%	85.71%	
	6	60%	99.06%	93.02%	100.00%	96.39%	
	7	80%	98.44%	90.70%	97.50%	93.98%	
	8	100%	98.75%	100.00%	90.00%	94.74%	
Test810	1	1%	98.75%	90.91%	100.00%	95.24%	87.50%
	2	5%	95.00%	75.00%	90.00%	81.82%	
	3	10%	92.50%	70.00%	70.00%	70.00%	
	4	20%	93.75%	77.78%	70.00%	73.68%	
	5	40%	97.50%	100.00%	80.00%	88.89%	
	6	60%	100.00%	100.00%	100.00%	100.00%	
	7	80%	98.75%	90.91%	100.00%	95.24%	
	8	100%	98.75%	100.00%	90.00%	94.74%	
Train830	1	1%	99.17%	93.75%	100.00%	96.77%	83.75%
	2	5%	95.83%	77.78%	93.33%	84.85%	
	3	10%	92.50%	68.75%	73.33%	70.97%	
	4	20%	93.33%	76.92%	66.67%	71.43%	
	5	40%	97.08%	100.00%	76.67%	86.79%	
	6	60%	96.67%	78.95%	100.00%	88.24%	
	7	80%	95.00%	82.14%	76.67%	79.31%	
	8	100%	97.92%	100.00%	83.33%	90.91%	
Test820	1	1%	98.75%	90.91%	100.00%	95.24%	83.75%
	2	5%	95.00%	75.00%	90.00%	81.82%	
	3	10%	92.50%	70.00%	70.00%	70.00%	
	4	20%	93.75%	77.78%	70.00%	73.68%	
	5	40%	96.88%	100.00%	75.00%	85.71%	
	6	60%	96.88%	80.00%	100.00%	88.89%	
	7	80%	95.63%	84.21%	80.00%	82.05%	
	8	100%	98.13%	100.00%	85.00%	91.89%	

wheat stripe rust in the training set Train840 or the testing set Test810.

Under the condition without considering single healthy wheat leaves, when the modeling ratio was 3:2, by using the severity assessment model of wheat stripe rust built with the *K*-means clustering algorithm based on the training set Train830, the severity assessments of the specimens in the training set Train830 and the corresponding testing set Test820 were implemented. As shown in Table 2, for all the severity classes of wheat stripe rust in the training set Train830, the lowest accuracy was 92.50% and the highest accuracy was 99.17%; the lowest precision was 68.75% and the highest precision was 100.00%; the lowest recall was 66.67%, and the highest recall was 100.00%; and the lowest F1 score was 70.97%, and the highest F1 score was 96.77%. The results showed

TABLE 3 Severity assessment results of the single diseased wheat leaves with the actual percentages of lesion areas of all the severity classes of wheat stripe rust contained in the training and testing sets, obtained by using the severity assessment models of wheat stripe rust built based on spectral clustering under the condition without considering single healthy wheat leaves.

Dataset	Severity category	Severity class	Accuracy	Precision	Recall	F1 score	Overall accuracy
Train840	1	1%	100.00%	100.00%	100.00%	100.00%	97.50%
	2	5%	98.44%	88.89%	100.00%	94.12%	
	3	10%	98.44%	100.00%	87.50%	93.33%	
	4	20%	100.00%	100.00%	100.00%	100.00%	
	5	40%	100.00%	100.00%	100.00%	100.00%	
	6	60%	100.00%	100.00%	100.00%	100.00%	
	7	80%	99.06%	93.02%	100.00%	96.39%	
	8	100%	99.06%	100.00%	92.50%	96.10%	
Test810	1	1%	100.00%	100.00%	100.00%	100.00%	98.75%
	2	5%	98.75%	90.91%	100.00%	95.24%	
	3	10%	98.75%	100.00%	90.00%	94.74%	
	4	20%	100.00%	100.00%	100.00%	100.00%	
	5	40%	100.00%	100.00%	100.00%	100.00%	
	6	60%	100.00%	100.00%	100.00%	100.00%	
	7	80%	100.00%	100.00%	100.00%	100.00%	
	8	100%	100.00%	100.00%	100.00%	100.00%	
Train830	1	1%	100.00%	100.00%	100.00%	100.00%	97.50%
	2	5%	98.75%	90.91%	100.00%	95.24%	
	3	10%	98.75%	100.00%	90.00%	94.74%	
	4	20%	100.00%	100.00%	100.00%	100.00%	
	5	40%	100.00%	100.00%	100.00%	100.00%	
	6	60%	100.00%	100.00%	100.00%	100.00%	
	7	80%	98.75%	90.91%	100.00%	95.24%	
	8	100%	98.75%	100.00%	90.00%	94.74%	
Test820	1	1%	100.00%	100.00%	100.00%	100.00%	96.25%
	2	5%	98.13%	86.96%	100.00%	93.02%	
	3	10%	98.13%	100.00%	85.00%	91.89%	
	4	20%	100.00%	100.00%	100.00%	100.00%	
	5	40%	99.38%	100.00%	95.00%	97.44%	
	6	60%	99.38%	95.24%	100.00%	97.56%	
	7	80%	98.75%	90.91%	100.00%	95.24%	
	8	100%	98.75%	100.00%	90.00%	94.74%	

that, for all the severity classes of wheat stripe rust in the testing set Test820, the lowest and highest accuracies were 92.50% and 98.75%, respectively; the lowest and highest precisions were 70.00% and 100.00%, respectively; the lowest and highest recalls were 70.00% and 100.00%, respectively; and the lowest and highest F1 scores were 70.00% and 95.24%, respectively. In terms of precision, recall, and F1 score, there were great differences between the severity assessment results for each severity class in the training set Train830 or the testing set Test820 achieved by using the model built based on the *K*-means clustering algorithm under the condition without considering single healthy wheat leaves. For the built severity assessment model, the overall accuracies of the training set Train830 and the corresponding testing set Test820 were both 83.75%.

Under the condition without considering single healthy wheat leaves, when the modeling ratio was 3:2, the severity assessment model of wheat stripe rust built based on the training set Train830 by using spectral clustering was used to carry out the severity assessments of the specimens in the training set Train830 and the corresponding testing set Test820. As shown in Table 3, for the training set Train830, except that the accuracies for the severity classes of 5%, 10%, 80%, and 100% were all 98.75%, the accuracies for all other severity classes were 100.00%; except that the precisions for the severity classes of 5% and 80% were both 90.91%, the precisions for all other severity classes were 100.00%; except that the recalls for the severity classes of 10% and 100% were both 90.00%, the recalls for all other severity classes were 100.00%; and among the F1 scores for all the severity classes, the lowest and highest F1 values were 94.74% and 100.00%, respectively. The results showed that, for all the severity classes of wheat stripe rust in the testing set Test820, the lowest and highest accuracies were 98.13% and 100.00%, respectively; the lowest and highest precisions were 86.96% and 100.00%, respectively; except that the recalls for the severity classes of 10%, 40%, and 100% were 85.00%, 95.00%, and 90.00%, respectively, the recalls for other severity classes were all 100.00%; and the lowest and highest F1 scores were 91.89% and 100.00%, respectively. For the built severity assessment model of wheat stripe rust based on spectral clustering, the overall accuracies of the training set Train830 and the corresponding testing set Test820 were 97.50% and 96.25%, respectively. The results indicated that, under the condition without considering single healthy wheat leaves, very good severity assessment results of the training and testing sets by using the severity assessment model built based on spectral clustering when the modeling ratio was 3:2, were achieved, and there were relatively small differences among the severity assessment performances of the built model on all the severity classes in the training set Train830 used for modeling or the corresponding testing set Test820.

The results indicated that, when the severity assessments of the single diseased leaf image datasets of wheat stripe rust without single healthy wheat leaves were conducted by using the severity assessment model built based on the two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering, respectively, the acceptable assessment results could be achieved. Under the two conditions with the modeling ratios equal to 4:1 and 3:2, there was little difference between the assessment performances obtained by using the severity assessment models of wheat stripe rust built based on each of the two unsupervised learning methods. However, the assessment performance obtained by using the severity assessment model built when the modeling ratio was 4:1, was slightly better than that obtained by using the severity assessment model built when the modeling ratio was 3:2. In the case of any modeling ratio, the assessment performance obtained by using the severity assessment model of wheat stripe rust based on spectral clustering was better than that obtained by using the severity assessment model of wheat stripe rust based on the *K*-means clustering algorithm, and the severity assessment model of wheat stripe rust based on spectral clustering could achieve very ideal assessment results on both the training set and the corresponding testing set, indicating that the severity assessment model of wheat stripe rust based on spectral clustering could be used as the optimal model for severity assessment of wheat stripe rust. Therefore, to achieve ideal severity assessment results by using the severity assessment method of wheat stripe rust based on unsupervised learning, the spectral clustering method can be used to build the severity assessment model of the disease. The results indicated that the wheat stripe rust severity assessment methods based on unsupervised learning proposed in this study could be applied to severity assessment of the disease under the condition without considering single healthy wheat leaves.

3.3 Severity assessment results obtained by using the severity assessment models of wheat stripe rust built based on the three supervised learning methods including the SVM, RF, and KNN under the condition without considering single healthy wheat leaves

Under the condition without considering single healthy wheat leaves, when the modeling ratio was 4:1, based on the training set Train840, the optimal SVM model for severity assessment of wheat stripe rust was built with the optimal parameter *C* equal to 1.741 and the optimal parameter *g* equal to 6.964; the optimal RF model for severity assessment of wheat stripe rust was built with the optimal number of decision trees equal to 10; and the optimal KNN model for severity assessment of wheat stripe rust was built with the optimal *K* of 3. Under the condition without considering single healthy wheat leaves, when the modeling ratio was 3:2, based on the training set Train830, the optimal SVM model for severity assessment of wheat stripe rust was built with the optimal parameter *C* of 0.758 and the optimal parameter *g* of 6.964; the optimal RF model for severity assessment of wheat stripe rust was built with the optimal number of decision trees equal to 20; and the optimal KNN model for severity assessment of wheat stripe rust was built with the optimal *K* equal to 5.

Under the condition without considering single healthy wheat leaves, when the modeling ratios were 4:1 and 3:2, the optimal SVM models and the optimal RF models built for severity assessment of wheat stripe rust were used to carry out the severity assessments of the specimens in the training sets (Train840 and Train830) and the testing sets (Test810 and Test820). For all the severity classes of wheat stripe rust in the training sets (Train840 and Train830) and the testing sets (Test810 and Test820), the accuracies, precisions, recalls, and F1 scores were all 100.00%. For each modeling ratio, by using the optimal SVM model and the optimal RF model built for severity assessment of wheat stripe rust, the overall accuracies of the training set that was used for modeling and the corresponding testing set were both 100.00% (Table 4).

Under the condition without considering single healthy wheat leaves, when the modeling ratio were 4:1, by using the optimal KNN model to perform the severity assessments of the specimens in the training set Train840 for modeling and the corresponding testing set Test810, the accuracy, precision, recall, and F1 score for the severity class of 1% in the training set Train840 were 99.69%, 100.00%, 97.50%, and 98.73%, respectively; the accuracy, precision, recall, and F1 score for the severity class of 5% in the training set Train840 were 99.69%, 97.56%, 100.00%, and 98.77%, respectively; and the accuracies, precisions, recalls,

TABLE 4 Overall accuracies of the training and testing sets obtained by using the built severity assessment SVM, RF, and KNN models of wheat stripe rust under the condition without considering single healthy wheat leaves.

Modeling method	Modeling ratio	Overall accuracy of the training set	Overall accuracy of the testing set
SVM	4:1	100.00%	100.00%
	3:2	100.00%	100.00%
RF	4:1	100.00%	100.00%
	3:2	100.00%	100.00%
KNN	4:1	99.69%	100.00%
	3:2	100.00%	100.00%

The table shows only the assessment results of the optimal severity assessment SVM, RF, and KNN models of wheat stripe rust.

and F1 scores for all other severity classes in the training set Train840 and all the severity classes in the testing set Test810 were 100.00%. Under the condition without considering single healthy wheat leaves, when the modeling ratios was 3:2, by using the optimal KNN model built based on the training set Train830 for severity assessment of wheat stripe rust, the accuracies, precisions, recalls, and F1 scores were all 100.00% for all the severity classes of wheat stripe rust in the training set Train830 and the testing set Test820. As shown in Table 4, for the modeling ratio of 4:1, by using the optimal KNN model built for severity assessment of wheat stripe rust based on the training set Train840, the overall accuracies of the training set Train840 and the corresponding testing set Test810 were 99.69% and 100.00%, respectively. For the modeling ratio of 3:2, by using the optimal KNN model built for severity assessment of wheat stripe rust based on the training set Train830, the overall accuracies of the training set Train830 and the corresponding testing set Test820 were both 100.00%.

The results showed that, under the condition without considering single healthy wheat leaves, by using the optimal SVM, RF, and KNN models for severity assessment of wheat stripe rust, very good severity assessment performances on the training sets (Train840 and Train830) and testing sets (Test810 and Test820) were achieved. In comparison, in the case of the two modeling ratios of 4:1 and 3:2, among the three kinds of models, the severity assessment models of wheat stripe rust built based on SVM and RF were optimal, and the corresponding overall accuracies of the training sets and the testing sets reached 100.00%. The results indicated that the methods for severity assessment of wheat stripe rust based on supervised learning proposed in this study could be applied to severity assessment of the disease under the condition without considering single healthy wheat leaves.

3.4 Severity assessment results obtained by using the severity assessment models of wheat stripe rust built based on the two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering under the condition of considering single healthy wheat leaves

Under the condition of considering single healthy wheat leaves, the severity assessment models of wheat stripe rust, built based on the two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering, were used to perform the severity assessments of the specimens of all the severity classes in the training and testing sets. The severity assessment results obtained by using the two unsupervised learning methods are shown in Tables 5, 6, respectively.

Under the condition of considering single healthy wheat leaves, by using the severity assessment model of wheat stripe rust built with the *K*-means clustering algorithm based on the training set Train940 when the modeling ratio was 4:1, the obtained assessment results of the training set Train940 as shown in Table 5, demonstrated that, for all the severity classes of the disease, high accuracies with the lowest value equal to 93.89% and the highest value equal to 99.17%, were achieved, but there were large differences in precision, recall, and F1 score. For all the severity classes of the disease in the training set Train940, the lowest and highest precisions were 71.43% and 100.00%, respectively; the lowest and highest recalls were 70.00% and 100.00%, respectively; and the lowest and highest F1 scores were 73.17% and 96.39%, respectively. The overall accuracy of the training set Train940 achieved by using the model was 84.72%. By using the built model to conduct the severity assessments of the specimens in the testing set Test910, for all the severity classes of wheat stripe rust, high accuracies were achieved, with the lowest value equal to 93.33% and the highest value equal to 100.00%; there were also large differences in precision, recall, and F1 score. The lowest precision, recall, and F1 score were all 70.00%, and the highest precision, recall, and F1 score were all 100.00%. For the built severity assessment model, the overall accuracy of the testing set Test910 was 85.56%.

As shown in Table 6, under the condition of considering single healthy wheat leaves, by using the severity assessment model of wheat stripe rust built with spectral clustering based on the training set Train940 when the modeling ratio was 4:1, the obtained assessment results for all the severity classes of wheat stripe rust in the training set Train940 showed that the lowest and highest accuracies was 98.33% and 100.00%, respectively; except that the precisions for the severity classes of 0%, 5%, and 80% were 95.24%, 86.96%, and 93.02%, respectively, the precisions for all other severity classes were 100.00%; except that the recalls for the severity classes of 1%, 10% and 100% were 95.00%, 85.00%, and 92.50%, respectively, the recalls for all other severity classes were

10.3389/fpls.2023.1150855

TABLE 5 Severity assessment results of the single wheat leaves with the actual percentages of lesion areas of all the severity classes of wheat stripe rust contained in the training and testing sets, obtained by using the severity assessment models of wheat stripe rust built based on the *K*-means clustering algorithm under the condition of considering single healthy wheat leaves.

Dataset	Severity category	Severity class	Accuracy	Precision	Recall	F1 score	Overall accuracy
Train940	0	0%	96.67%	76.92%	100.00%	86.96%	84.72%
	1	1%	95.00%	82.35%	70.00%	75.68%	
	2	5%	95.56%	77.27%	85.00%	80.95%	
	3	10%	93.89%	71.43%	75.00%	73.17%	
	4	20%	94.44%	77.78%	70.00%	73.68%	
	5	40%	97.22%	100.00%	75.00%	85.71%	
	6	60%	99.17%	93.02%	100.00%	96.39%	
	7	80%	98.61%	90.70%	97.50%	93.98%	
	8	100%	98.89%	100.00%	90.00%	94.74%	
Test910	0	0%	96.67%	76.92%	100.00%	86.96%	85.56%
	1	1%	95.56%	87.50%	70.00%	77.78%	
	2	5%	95.56%	75.00%	90.00%	81.82%	
	3	10%	93.33%	70.00%	70.00%	70.00%	
	4	20%	94.44%	77.78%	70.00%	73.68%	
	5	40%	97.78%	100.00%	80.00%	88.89%	
	6	60%	100.00%	100.00%	100.00%	100.00%	
	7	80%	98.89%	90.91%	100.00%	95.24%	
	8	100%	98.89%	100.00%	90.00%	94.74%	
Train930	0	0%	96.67%	76.92%	100.00%	86.96%	85.19%
	1	1%	95.56%	87.50%	70.00%	77.78%	
	2	5%	95.93%	77.14%	90.00%	83.08%	
	3	10%	93.70%	70.97%	73.33%	72.13%	
	4	20%	94.44%	77.78%	70.00%	73.68%	
	5	40%	97.41%	100.00%	76.67%	86.79%	
	6	60%	99.26%	93.75%	100.00%	96.77%	
	7	80%	98.52%	90.63%	96.67%	93.55%	
	8	100%	98.89%	100.00%	90.00%	94.74%	
Test920	0	0%	96.67%	76.92%	100.00%	86.96%	85.56%
	1	1%	95.56%	87.50%	70.00%	77.78%	
	2	5%	96.11%	78.26%	90.00%	83.72%	
	3	10%	93.89%	71.43%	75.00%	73.17%	
	4	20%	94.44%	77.78%	70.00%	73.68%	
	5	40%	97.22%	100.00%	75.00%	85.71%	
	6	60%	99.44%	95.24%	100.00%	97.56%	
	7	80%	98.89%	90.91%	100.00%	95.24%	
	8	100%	98.89%	100.00%	90.00%	94.74%	

10.3389/fpls.2023.1150855

TABLE 6 Severity assessment results of the single wheat leaves with the actual percentages of lesion areas of all the severity classes of wheat stripe rust contained in the training and testing sets, obtained by using the severity assessment models of wheat stripe rust built based on spectral clustering under the condition of considering single healthy wheat leaves.

Dataset	Severity category	Severity class	Accuracy	Precision	Recall	F1 score	Overall accuracy
Train940	0	0%	99.44%	95.24%	100.00%	97.56%	96.94%
	1	1%	99.44%	100.00%	95.00%	97.44%	
	2	5%	98.33%	86.96%	100.00%	93.02%	
	3	10%	98.33%	100.00%	85.00%	91.89%	
	4	20%	100.00%	100.00%	100.00%	100.00%	
	5	40%	100.00%	100.00%	100.00%	100.00%	
	6	60%	100.00%	100.00%	100.00%	100.00%	
	7	80%	99.17%	93.02%	100.00%	96.39%	
	8	100%	99.17%	100.00%	92.50%	96.10%	
Test910	0	0%	98.89%	90.91%	100.00%	95.24%	97.78%
	1	1%	98.89%	100.00%	90.00%	94.74%	
	2	5%	98.89%	90.91%	100.00%	95.24%	
	3	10%	98.89%	100.00%	90.00%	94.74%	
	4	20%	100.00%	100.00%	100.00%	100.00%	
	5	40%	100.00%	100.00%	100.00%	100.00%	
	6	60%	100.00%	100.00%	100.00%	100.00%	
	7	80%	100.00%	100.00%	100.00%	100.00%	
	8	100%	100.00%	100.00%	100.00%	100.00%	
Train930	0	0%	99.26%	93.75%	100.00%	96.77%	96.67%
	1	1%	99.26%	100.00%	93.33%	96.55%	
	2	5%	98.52%	88.24%	100.00%	93.75%	
	3	10%	98.52%	100.00%	86.67%	92.86%	
	4	20%	100.00%	100.00%	100.00%	100.00%	
	5	40%	99.63%	100.00%	96.67%	98.31%	
	6	60%	99.63%	96.77%	100.00%	98.36%	
	7	80%	99.26%	93.75%	100.00%	96.77%	
	8	100%	99.26%	100.00%	93.33%	96.55%	
Test920	0	0%	99.44%	95.24%	100.00%	97.56%	95.56%
	1	1%	99.44%	100.00%	95.00%	97.44%	
	2	5%	98.33%	86.96%	100.00%	93.02%	
	3	10%	98.33%	100.00%	85.00%	91.89%	
	4	20%	99.44%	95.24%	100.00%	97.56%	
	5	40%	98.89%	100.00%	90.00%	94.74%	
	6	60%	99.44%	95.24%	100.00%	97.56%	
	7	80%	98.89%	90.91%	100.00%	95.24%	
	8	100%	98.89%	100.00%	90.00%	94.74%	

100.00%; and the lowest and highest F1 scores were 91.89% and 100.00%, respectively. By using the built model to conduct the severity assessments of the specimens in the testing set Test910, the obtained results showed that, except that the accuracies for the severity classes of 0%, 1%, 5%, and 10% were all 98.89%, the accuracies for other severity classes were all 100.00%; except that the precisions for the severity classes of 0% and 5% were both 90.91%, the precisions for other severity classes were all 100.00%; except that the recalls for the severity classes of 1% and 10% were both 90.00%, the recalls for other severity classes were all 100.00%; except that the F1 scores for the severity classes of 0%, 1%, 5%, and 10% were 95.24%, 94.74%, 95.24%, and 94.74%, respectively, the F1 scores for all other severity classes were 100.00%. For the severity assessment model of wheat stripe rust built with spectral clustering based on the training set Train940, the overall accuracies of the training set Train940 and the corresponding testing set Test910 were 96.94% and 97.78%, respectively. The results indicated that very good severity assessment results for the specimens in both the training set and the testing set were achieved by using the severity assessment model built with spectral clustering when the modeling ratio was 4:1 under the condition of considering single healthy wheat leaves.

Under the condition of considering single healthy wheat leaves, when the modeling ratio was 3:2, by using the severity assessment model of wheat stripe rust built with the *K*-means clustering algorithm based on the training set Train930, the severity assessment results (Table 5) showed that, for all the severity classes of wheat stripe rust in the training set Train930, high accuracies were obtained, with the lowest value equal to 93.70% and the highest value equal to 99.26%; the lowest and highest precisions were 70.97% and 100.00%, respectively; the lowest and highest recalls were 70.00% and 100.00%, respectively; and the lowest and highest F1 scores were 72.13% and 96.77%, respectively. The overall accuracy of the training set Train930 achieved by using the built severity assessment model was 85.19%. The results indicated that there were great differences in precisions, recalls, and F1 scores for all the severity classes in the training set Train930 achieved by using the model built based on the *K*-means clustering algorithm. The assessment results obtained by using the severity assessment model of wheat stripe rust built with the *K*-means clustering algorithm based on the training set Train930 showed that, for all the severity classes of wheat stripe rust in the testing set Train920, the lowest and highest accuracies were 93.89% and 99.44%, respectively; the lowest and highest precisions were 71.43% and 100.00%, respectively; the lowest and highest recalls were 70.00% and 100.00%, respectively; and the lowest and highest F1 scores were 73.17% and 97.56%, respectively. There were also great differences in precisions, recalls, and F1 scores for all the severity classes in the testing set Train920 achieved by using the model built based on the *K*-means clustering algorithm under the condition of considering single healthy wheat leaves. For the built severity assessment model, the overall accuracy of the testing set Test920 was 85.56%.

As shown in Table 6, under the condition of considering single healthy wheat leaves, by using the severity assessment model of wheat stripe rust built with spectral clustering based on the training set Train930 when the modeling ratio was 3:2, the obtained assessment results showed that, for all the severity classes of wheat stripe rust in the training set Train930, the lowest and highest accuracies were 98.52% and 100.00%, respectively; the lowest and highest precisions were 88.24% and 100.00%, respectively; except that the recalls for the severity classes of 1%, 10%, 40%, and 100% were 93.33%, 86.67%, 96.67%, and 93.33%, respectively, the recalls for other severity classes were all 100.00%; and the lowest F1 score was 92.86% and the highest F1 score was 100.00%. The assessment results obtained by using the severity assessment model of wheat stripe rust built with spectral clustering based on the training set Train930 showed that, for all the severity classes of wheat stripe rust in the testing set Test920, the lowest and highest accuracies were 98.33% and 99.44%, respectively; the lowest and highest precisions were 86.96% and 100.00%, respectively; except that the recalls for the severity classes of 1%, 10%, 40%, and 100% were 95.00%, 85.00%, 90.00%, and 90.00%, respectively, the recalls for all other severity classes were 100.00%; and the lowest F1 score was 91.89%, and the highest F1 score was 97.56%. For the severity assessment model of wheat stripe rust built with spectral clustering based on Train930, the overall accuracies of the training set Train930 and the corresponding testing set Test920 were 96.67% and 95.56%, respectively. The results indicated that, under the condition of considering single healthy wheat leaves, very good severity assessment performance on both the training set Train930 and the corresponding testing set Test920 were achieved by using the severity assessment model of wheat stripe rust built with spectral clustering.

The results showed that, when the severity assessments of the single wheat leaf image datasets containing single healthy wheat leaves were conducted by using the severity assessment models, built based on the two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering, respectively, the achieved assessment performances were acceptable. When the modeling ratio was 4:1 or 3:2, the assessment performance obtained by using the severity assessment model of wheat stripe rust based on spectral clustering was better than that obtained by using the severity assessment model of wheat stripe rust based on the *K*-means clustering algorithm, and very ideal assessment performances on both the training set and the corresponding testing set could be achieved by using the severity assessment model of wheat stripe rust based on spectral clustering, indicating that the severity assessment model of wheat stripe rust based on spectral clustering could be treated as the optimal model for carrying out severity assessment of wheat stripe rust. Therefore, to achieve ideal severity assessment results by using the severity assessment method of wheat stripe rust based on unsupervised learning under the condition of considering single healthy wheat leaves, the severity assessment model of the disease can be built based on spectral clustering. The results demonstrated that the methods for severity assessment of wheat stripe rust based on unsupervised learning proposed in this study could be utilized to carry out severity assessment of wheat stripe rust under the condition of considering single healthy wheat leaves.

3.5 Severity assessment results obtained by using the severity assessment models of wheat stripe rust built based on the three supervised learning methods including the SVM, RF, and KNN under the condition of considering single healthy wheat leaves

Under the condition of considering single healthy wheat leaves, when the modeling ratio was 4:1, based on the training set Train940, the optimal SVM model for severity assessment of wheat stripe rust was built with the optimal parameter *C* of 2.297 and the optimal parameter *g* of 4.000. By using this SVM model to carry out the severity assessments of the specimens in the training set Train940 and the testing set Test910, for the severity class of 0% (healthy wheat leaves) in the training set Train940, the accuracy, precision, recall, and F1 score were 99.72%, 97.56%, 100.00%, and 98.77%, respectively; for the severity class of 1% in the training set Train940, the accuracy, precision, recall, and F1 score were 99.72%, 100.00%, 97.50%, and 98.73%, respectively; and for all other severity classes in the training set Train940 and all the severity classes in the testing set Test910, the accuracies, precisions, recalls, and F1 scores were all 100.00%. For the built optimal SVM model for severity assessment of wheat stripe rust when the modeling ratio was 4:1 under the condition of considering single healthy wheat leaves, the overall accuracies of the training set Train940 and the corresponding testing set Test910 were 99.72% and 100.00%, respectively (Table 7).

Under the condition of considering single healthy wheat leaves, when the modeling ratio was 3:2, based on the training set Train930, the optimal SVM model for severity assessment of wheat stripe rust was built with the optimal parameter *C* equal to 194.012 and the optimal parameter *g* equal to 0.330. By using this SVM model to perform the severity assessments of the specimens in the training set Train930 that was used for modeling and the corresponding testing set Test920, for the severity class of 1% in the training set Train930, the accuracy, precision, recall, and F1 score were 99.63%, 100.00%, 96.67%, and 98.31%, respectively; for the severity class of 5% in the training set Train930, the accuracy, precision, recall, and F1 score were 99.63%, 96.77%, 100.00%, and 98.36%, respectively; and for all other severity classes in the training set Train930 and all the severity classes in the testing set Test920, the accuracies, precisions, recalls, and F1 scores were all 100.00%. For the built optimal SVM model for severity assessment of wheat stripe rust when the modeling ratio was 3:2 under the condition of considering single healthy wheat leaves, the overall accuracies of the training set Train930 and the corresponding testing set Test920 were 99.63% and 100.00%, respectively (Table 7).

Under the condition of considering single healthy wheat leaves, when the modeling ratio was 4:1, based on the training set Train940, the optimal RF model for severity assessment of wheat stripe rust was built with the optimal number of decision trees equal to 10; and when the modeling ratio was 3:2, based on the training set Train930, the optimal RF model for severity assessment of wheat stripe rust was built with the optimal number of decision trees equal to 20. Under the two conditions of the modeling ratios equal to 4:1 and 3:2, the optimal RF models built for severity assessment of wheat stripe rust were used to conduct the severity assessments of the specimens in the training sets (Train940 and Train930) and the testing sets (Test910 and Test920), and for all the severity classes of wheat stripe rust in the training sets and the testing sets, the accuracies, precisions, recalls, and F1 scores were all 100.00%. For each modeling ratio under the condition of considering single healthy wheat leaves, by using the optimal RF model built for severity assessment of wheat stripe rust, the overall accuracies of the training set that was used for modeling and the corresponding testing set were both 100.00% (Table 7).

Under the condition of considering single healthy wheat leaves, when the modeling ratio was 4:1, based on the training set Train940, the optimal KNN model for severity assessment of wheat stripe rust was built with the optimal *K* of 9. By using this KNN model to perform the severity assessments of the specimens in the training set Train940 and the testing set Test910, for the severity classes of 0%, 5%, and 80% in the training set Train940, the accuracies were all 99.72%, the precision were all 97.56%, the recall were all 100.00%, and F1 score were all 98.77%; for the severity classes of 1%, 10%, and 100% in the training set Train940, the accuracies were all 99.72%, the precisions were all 100.00%, the recalls were all 97.50%, and the F1 scores were all 98.73%; and for all other severity classes in the training set Train940 and all the severity classes in the testing set Test910, all the accuracies, precisions, recalls, and F1 scores were 100.00%. As shown in Table 7, for the built optimal KNN model for severity assessment of wheat stripe rust when the modeling ratio was 4:1 under the condition of considering single healthy wheat leaves, the overall accuracy of the training set Train940 was 99.17%, and that of the corresponding testing set Test910 was 100.00%.

TABLE 7 Overall accuracies of the training and testing sets obtained by using the built severity assessment SVM, RF, and KNN models of wheat stripe rust under the condition of considering single healthy wheat leaves.

Modeling method	Modeling ratio	Overall accuracy of the training set	Overall accuracy of the testing set
SVM	4:1	99.72%	100.00%
	3:2	99.63%	100.00%
RF	4:1	100.00%	100.00%
	3:2	100.00%	100.00%
KNN	4:1	99.17%	100.00%
	3:2	99.63%	100.00%

The table shows only the assessment results of the optimal severity assessment SVM, RF, and KNN models of wheat stripe rust.

Under the condition of considering single healthy wheat leaves, when the modeling ratio was 3:2, based on the training set Train930, the optimal KNN model for severity assessment of wheat stripe rust was built with the optimal *K* equal to 5. By using this optimal KNN model to perform the severity assessments of the specimens in the training set Train930 and the testing set Test920, for the severity class of 0% in the training set Train930, the accuracy, precision, recall, and F1 score were 99.63%, 96.77%, 100.00%, and 98.36%, respectively; for the severity class of 1% in the training set Train930, the accuracy, precision, recall, and F1 score were 99.63%, 100.00%, 96.67%, and 98.31%, respectively; and for all other severity classes in the training set Train930 and all the severity classes in the testing set Test920, all the accuracies, precisions, recalls, and F1 scores were 100.00%. As shown in Table 7, for the built optimal KNN model for severity assessment of wheat stripe rust when the modeling ratio was 3:2 under the condition of considering single healthy wheat leaves, the overall accuracies of the training set Train930 and the corresponding testing set Test920 were 99.63% and 100.00%, respectively.

The results indicated that, under the condition of considering single healthy wheat leaves, very good severity assessment performances on the training sets (Train940 and Train930) and testing sets (Test910 and Test920) could be obtained by using the built optimal SVM, RF, and KNN models for severity assessment of wheat stripe rust. Furthermore, in the case of the two modeling ratios of 4:1 and 3:2, the severity assessment performances of the built optimal RF models were the best, and the overall accuracies of the training sets and the testing sets were all 100.00%. The obtained results indicated that the severity assessment methods based on supervised learning for wheat stripe rust proposed in this study could be used to carry out severity assessment of wheat stripe rust under the condition of considering single healthy wheat leaves.

4 Discussion

In this study, the methods for severity assessment of wheat stripe rust were proposed based on machine learning. Regardless of whether the healthy wheat leaves were considered or not, acceptable assessment performances could be obtained by using the severity assessment models of wheat stripe rust, built with the two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering, based on the training sets and the testing sets constructed by using the system sampling method with the modeling ratios of 4:1 and 3:2. Especially, high accuracies, precisions, recalls, F1 scores, and overall accuracies were obtained on both the training sets and the testing sets by using the severity assessment models of wheat stripe rust built based on spectral clustering, and the relatively ideal performances for severity assessments of wheat stripe rust were achieved. Regardless of whether the healthy wheat leaves were considered or not, very good assessment results were achieved by using the severity assessment models of wheat stripe rust built based on the three supervised learning methods including SVM, RF, and KNN when the modeling ratio were 4:1 and 3:2. In particular, the accuracies, precisions, recalls, and F1 scores for all the severity classes of the training and testing sets and the overall accuracies of the training and testing sets, were all 100.00%, by using the optimal models (the optimal RF models) for severity assessment of wheat stripe rust obtained by comparing the three modeling methods of SVM, RF and KNN, and thus, very ideal performances for severity assessments of wheat stripe rust were achieved by using the selected optimal models (the optimal RF models). The results indicated that the methods for wheat stripe rust severity assessment based on unsupervised learning and supervised learning proposed in this study could be used for severity assessment of wheat stripe rust. For the proposed severity assessment methods of wheat stripe rust in this study, the model building and severity assessments were conducted based on the actual percentages of lesion areas in the areas of the corresponding whole single diseased wheat leaves. The problem that, when severity assessment of wheat stripe rust is carried out based on the percentage of the lesion area in the area of a whole single diseased wheat leaf, the percentage of the lesion area in the area of the whole single diseased wheat leaf corresponding to a severity class in the severity grading standard of the disease is not inconsistent with the actual percentage of the lesion area in the area of the whole single diseased wheat leaf, was completely solved in this study. By using the methods proposed in this study, based on the actual percentage of the lesion area in the whole area of a single diseased wheat leaf, the severity of the corresponding diseased leaf can be directly assessed. The results obtained in this study provide a basis for accurately assessing the severity of wheat stripe rust. The methods and ideas provided in this study are also applicable to other plant diseases such as wheat leaf rust caused by *Puccinia triticina*, for which the percentage of the lesion area in the corresponding diseased plant unit area of a severity class in the corresponding disease severity grading standard is not inconsistent with the actual percentage of the lesion area in the area of the whole diseased plant unit. They are also applicable to other plant diseases for which disease severity assessment of a diseased plant unit is carried out based on the ratio of the lesion area to the area of the whole diseased plant unit. They can be used to solve such problems in the severity assessments of plant diseases. In this study, two unsupervised learning methods including the *K*-means clustering algorithm and spectral clustering and three supervised learning methods including SVM, RF, and KNN were used to build the severity assessment models of wheat stripe rust, respectively. By using the ideas provided in this study, other unsupervised learning methods and supervised learning methods can be used to build the severity assessment models of wheat stripe rust in further studies.

The severity assessment methods of wheat stripe rust proposed in this study were to build the severity assessment models of wheat stripe rust based on the actual percentages of lesion areas in the areas of the corresponding whole single diseased wheat leaves by using the unsupervised learning methods and the supervised learning methods. Then, by using the built severity assessment models of wheat stripe rust, the severity classes of the single diseased wheat leaves with the actual percentages of lesion areas could be directly assessed. In particular, when an unsupervised learning method is used to build the severity assessment model of wheat stripe rust, there is no need to artificially determine the severity

classes of the single diseased wheat leaves with the actual percentages of lesion areas, the single diseased wheat leaves can be classified and assessed by using the unsupervised learning method according to the number of severity classes in the disease severity grading standard, and then, through optimization of the built severity assessment model, a model for severity assessment of wheat stripe rust with satisfactory assessment performance can be obtained. Unsupervised learning is conducive to the application of the developed severity assessment methods in practice, and can reduce the errors in severity assessments caused by using the visual observation method. The severity assessment methods of wheat stripe rust based on machine learning developed in this study has strong practical applicability and can realize the accurate severity assessment of the disease, which is of great significance for the survey, monitoring, prediction, and control of the disease. After the severity classes of wheat leaves are assessed, disease prevalence evaluation and disease predication can be performed, and then suitable control measures can be made. Generally, the evaluated disease prevalence data or the disease prediction results are compared to the control threshold or economic threshold of the disease to determine whether the disease needs to be controlled. Once the control threshold or economic threshold is reached, suitable measures, such as spraying fungicides, can be taken to control the disease.

At the present time, severity assessments of wheat stripe rust based on image processing technology are mainly realized by comparing the actual percentage of the lesion area in the area of a whole single diseased wheat leaf obtained by using image processing to the percentages for the eight severity classes in the severity grading standard of the disease (Jiang et al., 2021) or by building the severity assessment models of wheat stripe rust based on extracted features of disease images (Bao et al., 2021). The percentage of the lesion area in the area of a whole diseased wheat leaf corresponding to each severity class in the severity grading standard of wheat stripe rust is obviously greater than the actual percentage of the lesion area in the area of the whole diseased wheat leaf, which may cause great errors in the disease severity assessments. Jiang et al. (2022) determined the reference ranges for severity assessment of wheat stripe rust based on the actual percentages of lesion areas corresponding to each severity class of wheat stripe rust, and then obtained satisfactory assessment results (the severity assessment accuracies for the training sets and testing sets were not lower than 85%.) of the diseased wheat leaves with the actual percentages of lesion areas according to the determined reference ranges. On the basis of the study conducted by Jiang et al. (2022), the severity assessment models of wheat stripe rust were built based on machine learning with the actual percentages of the lesion areas in the areas of the corresponding single leaves of wheat stripe rust and the corresponding severity category data in this study. According to the calculating methods, the two indicators used to evaluate the performances of severity assessment methods, recall used in this study and severity assessment accuracy used by Jiang et al. (2022), are the same. In terms of recall used in this study and severity assessment accuracy used by Jiang et al. (2022), a comparison of the performances of the methods developed in this study to those of the methods developed by Jiang et al. (2022) was made, and the results showed that the performance of the severity assessment method based on the *K*-means clustering algorithm developed in this study was the worst and that of the severity assessment method based on RF developed in this study was the best (Tables 8, 9). The severity assessment methods of wheat stripe rust based on machine learning developed in this study will greatly improve the accuracy of image-based severity assessment of the disease, and can greatly reduce the requirements for the experience of assessors/raters in disease severity assessment. This is beneficial to improve the reliability of the monitoring and early warning information of wheat stripe rust, and is conducive to the popularization and application of related technologies. This will be helpful to improve the level of the survey, monitoring and early warning, and management of wheat stripe rust. The automatization and intellectualization of plant disease severity assessment is an inevitable development trend of the science and technology (Wang, 2022). This study provides ideas and basis for accurate severity assessment of wheat stripe rust based on image processing technology, which is conducive to the development of automatic severity assessment system of the disease and the improvement of the severity assessment level of the disease. This is useful to promoting the automatization and intellectualization of severity assessment of wheat stripe rust, and can provide more reliable support for the prediction, variety resistance identification and variety breeding, and disease control strategy making of wheat stripe rust.

In this study, the actual percentages of lesion areas in the areas of the corresponding whole single diseased leaves infected by wheat stripe rust were obtained by using image processing technology with image processing software. The actual percentage data can be obtained by assessors/raters using the visual observation method. In addition, the actual percentages of lesion areas can be obtained by programming to implement the segmentation of lesion images and the calculation of ratios of the lesion areas to the areas of the corresponding whole single diseased leaves (Li et al., 2011; Jiang et al., 2021), by using special software and packages (Lamari, 2008; Schneider et al., 2012; Pethybridge and Nelson, 2015; Olivoto et al., 2022), and by actual experimental operation methods such as graph paper method and paper-weighing method (Li et al., 2011). The automatic lesion segmentation methods and the automatic calculation methods of the actual percentages of lesion areas in the areas of the corresponding whole single diseased leaves can be combined with the severity assessment methods and models developed in this study to construct automatic severity assessment systems of wheat stripe rust, which is conducive to the implementation of automatic severity assessment of wheat stripe rust and the more convenient practical applications of related technical methods.

By using the reference-range-based methods proposed by Jiang et al. (2022) to assess the severity class of a single diseased wheat leaf infected with wheat stripe rust, the actual percentage of the lesion area in the area of the whole single diseased leaf needs to be compared to the upper and lower limits of the reference ranges of all the severity classes of the disease. For the severity assessment methods of wheat stripe rust based on machine learning proposed in this study, the actual percentage of the lesion area in the area of a

TABLE 8 Comparison results of the performances of the machine-learning-based methods developed in this study to those of the reference-range-based methods developed by Jiang et al. (2022) when the modeling/sampling ratio was 4:1, in terms of recall used in this study and severity assessment accuracy used by Jiang et al. (2022) that are the same according to their calculating methods.

Dataset	Severity class	Recall used in this study										Severity assessment accuracy used by Jiang et al. (2022)			
		K-means clustering algorithm	Spectral clustering	SVM	RF	KNN	*K*-means clustering algorithm	Spectral clustering	SVM	RF	KNN	The mid-point-of-two-adja-cent-means-based actual percentage reference range	The 90% reference range	The 95% reference range	The 99% reference range
Training set	0%	–	–	–	–	–	100.00%	100.00%	100.00%	100.00%	100.00%	–	–	–	–
	1%	100.00%	100.00%	100.00%	100.00%	97.50%	70.00%	95.00%	97.50%	100.00%	97.50%	100.00%	95.00%	97.50%	100.00%
	5%	95.00%	100.00%	100.00%	100.00%	100.00%	85.00%	100.00%	100.00%	100.00%	100.00%	100.00%	95.00%	100.00%	100.00%
	10%	75.00%	87.50%	100.00%	100.00%	100.00%	75.00%	85.00%	100.00%	100.00%	97.50%	87.50%	90.00%	100.00%	100.00%
	20%	70.00%	100.00%	100.00%	100.00%	100.00%	70.00%	100.00%	100.00%	100.00%	100.00%	97.50%	95.00%	100.00%	100.00%
	40%	75.00%	100.00%	100.00%	100.00%	100.00%	75.00%	100.00%	100.00%	100.00%	100.00%	85.00%	95.00%	95.00%	100.00%
	60%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	92.50%	97.50%	95.00%
	80%	97.50%	100.00%	100.00%	100.00%	100.00%	97.50%	100.00%	100.00%	100.00%	100.00%	97.50%	90.00%	97.50%	100.00%
	100%	90.00%	92.50%	100.00%	100.00%	100.00%	90.00%	92.50%	100.00%	100.00%	97.50%	90.00%	87.50%	100.00%	100.00%
Testing set	0%	–	–	–	–	–	100.00%	100.00%	100.00%	100.00%	100.00%	–	–	–	–
	1%	100.00%	100.00%	100.00%	100.00%	100.00%	70.00%	90.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%
	5%	90.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%
	10%	70.00%	90.00%	100.00%	100.00%	100.00%	70.00%	90.00%	100.00%	100.00%	100.00%	90.00%	90.00%	100.00%	100.00%
	20%	70.00%	100.00%	100.00%	100.00%	100.00%	70.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	40%	80.00%	100.00%	100.00%	100.00%	100.00%	80.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%	100.00%
	60%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%
	80%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%
	100%	90.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%	100.00%	100.00%	90.00%	90.00%	100.00%	100.00%

'–' in the table denotes that single healthy wheat leaves were not considered.

10.3389/fpls.2023.1150855

TABLE 9 Comparison results of the performances of the machine-learning-based methods developed in this study to those of the reference-range-based methods developed by Jiang et al. (2022) when the modeling/sampling ratio was 3:2, in terms of recall used in this study and severity assessment accuracy used by Jiang et al. (2022) that are the same according to their calculating methods.

Dataset	Severity class	Recall used in this study										Severity assessment accuracy used by Jiang et al. (2022)			
		K-means clustering algorithm	Spectral clustering	SVM	RF	KNN	*K*-means clustering algorithm	Spectral clustering	SVM	RF	KNN	The mid-point-of-two-adja-cent-means-based actual percentage reference range	The 90% reference range	The 95% reference range	The 99% reference range
Training set	0%	–	–	–	–	–	100.00%	100.00%	100.00%	100.00%	100.00%	–	–	–	–
	1%	100.00%	100.00%	100.00%	100.00%	100.00%	70.00%	93.33%	96.67%	100.00%	96.67%	100.00%	93.33%	96.67%	100.00%
	5%	93.33%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%	100.00%	100.00%	100.00%	93.33%	100.00%	100.00%
	10%	73.33%	90.00%	100.00%	100.00%	100.00%	73.33%	86.67%	100.00%	100.00%	100.00%	86.67%	90.00%	96.67%	96.67%
	20%	66.67%	100.00%	100.00%	100.00%	100.00%	70.00%	100.00%	100.00%	100.00%	100.00%	96.67%	96.67%	100.00%	100.00%
	40%	76.67%	100.00%	100.00%	100.00%	100.00%	76.67%	96.67%	100.00%	100.00%	100.00%	86.67%	96.67%	96.67%	96.67%
	60%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	93.33%	96.67%	100.00%
	80%	76.67%	100.00%	100.00%	100.00%	100.00%	96.67%	100.00%	100.00%	100.00%	100.00%	96.67%	93.33%	96.67%	100.00%
	100%	83.33%	90.00%	100.00%	100.00%	100.00%	90.00%	93.33%	100.00%	100.00%	100.00%	90.00%	90.00%	100.00%	100.00%
Testing set	0%	–	–	–	–	–	100.00%	100.00%	100.00%	100.00%	100.00%	–	–	–	–
	1%	100.00%	100.00%	100.00%	100.00%	100.00%	70.00%	95.00%	100.00%	100.00%	100.00%	100.00%	95.00%	100.00%	100.00%
	5%	90.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%	100.00%	100.00%	100.00%	95.00%	100.00%	100.00%
	10%	70.00%	85.00%	100.00%	100.00%	100.00%	75.00%	85.00%	100.00%	100.00%	100.00%	90.00%	90.00%	95.00%	95.00%
	20%	70.00%	100.00%	100.00%	100.00%	100.00%	70.00%	100.00%	100.00%	100.00%	100.00%	100.00%	95.00%	100.00%	100.00%
	40%	75.00%	95.00%	100.00%	100.00%	100.00%	75.00%	90.00%	100.00%	100.00%	100.00%	85.00%	95.00%	95.00%	100.00%
	60%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%
	80%	80.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	90.00%	100.00%	100.00%
	100%	85.00%	90.00%	100.00%	100.00%	100.00%	90.00%	90.00%	100.00%	100.00%	100.00%	90.00%	85.00%	100.00%	100.00%

'–' in the table denotes that single healthy wheat leaves were not considered.

10.3389/fpls.2023.1150855

whole single diseased wheat leaf infected with wheat stripe rust needs to be input into the built severity assessment models, and then, the severity class of the single diseased wheat leaf to be assessed can be determined. The severity assessment methods of wheat stripe rust proposed in this study are different from those proposed by Jiang et al. (2022), however, by using all these methods, satisfactory assessment results can be achieved. Both the severity assessment methods developed in this study and those proposed by Jiang et al. (2022) can be automated by computer programming. According to the actual situation, these methods can be integrated with disease image processing systems, and the severity assessment functions of these systems can be improved or the severity assessment functions can be added to these systems, to realize the automation of severity assessment of wheat stripe rust.

In the surveys and assessments of wheat stripe rust, severity and infection type are two different terms. Generally, infection type is used as an important indicator to evaluate the resistance of wheat to stripe rust. The same strain or physiological race of *Pst* can cause different infection types on different wheat varieties. Determination of infection type is a kind of qualitative evaluation, which can be implemented by using extracted image features to identify the categories of infection types based on image processing technology (Hayit et al., 2021). Different severity classes of wheat stripe rust can also be qualitatively identified based on image processing technology (Bao et al., 2021). In fact, the actual percentage of lesion area is a continuous variable. Therefore, the disease severity assessment method based on the actual percentages of the lesion areas should be the best solution to implement severity assessment of wheat stripe rust.

5 Conclusions

In this study, efforts were made to solve the problems existing in the severity assessment methods of wheat stripe rust and to develop new methods for severity assessment of the disease based on machine learning. The acquired actual percentage data of the lesion areas of single diseased wheat leaves were used to construct the training sets and the corresponding testing sets by using the system sample method with two modeling ratios under the two conditions of considering healthy single wheat leaves or not, and then, the two unsupervised learning methods including *K*-means clustering algorithm and spectral clustering and the three supervised learning methods including SVM, RF, and KNN were used to build the severity assessment models of wheat stripe rust, respectively. By using the built models to carry out the severity assessments of the training and testing sets, satisfactory assessment results were achieved. In particular, the severity classes of all the specimens in the training and testing sets can correctly assessed by using the built optimal RF models for severity assessment of wheat stripe rust. The results indicated that good assessment performance can be achieved by using the disease severity assessment methods developed in this study, and that the methods are suitable for the severity assessment of wheat stripe rust. The severity assessment methods of wheat stripe rust based on machine learning were provided in this study. For the methods, the obtained actual percentages of lesion areas in the areas of the whole single diseased wheat leaves were directly used to build the severity assessment models of wheat stripe rust, and then, the built models were used to conduct severity assessments of single wheat leaves based on the obtained corresponding actual percentages of lesion areas of the leaves. The methods are simple, rapid, and easy-to-operate, and by using these methods, very highly accurate assessment results of single wheat leaves can be achieved. More importantly, the severity assessment methods proposed in this study can provide a reference for the severity assessments of all the plant diseases for which the severity assessments are performed based on the actual ratios of lesion areas to the areas of the diseased plant units, and this study can provide a basis for the implementation of automatic severity assessments of plant diseases by computer programming based on computer vision technology and image processing technology. In future studies, automatic severity assessment systems of plant diseases can be developed based on the proposed severity assessment methods. It is feasible for UAV (unmanned aerial vehicle) applications in field environments under the conditions that actual ratios of lesion areas of plant units could be obtained.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

HGW contributed conception of the study and designed the experiments. QJ, HLW and HGW performed the experiments; QJ, HLW and HGW analyzed the data; QJ and HGW wrote the draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (Grant No. 32072357) and the National Key Basic Research Program of China (Grant No. 2021YFD1401001 and 2018YFD0200402).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ali, S., Rodriguez-Algaba, J., Thach, T., Sørensen, C. K., Hansen, J. G., Lassen, P., et al. (2017). Yellow rust epidemics worldwide were caused by pathogen races from divergent genetic lineages. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01057

Bao, W. X., Lin, Z., Hu, G. S., Liang, D., Huang, L. S., and Yang, X. J. (2021). Severity estimation of wheat leaf diseases based on RSTCNN. *Trans. Chin. Soc Agric. Mach.* 52, 242–252, 263. doi: 10.6041/j.issn.1000-1298.2021.12.026

Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60. doi: 10.1016/j.biosystemseng.2016.01.017

Bock, C. H., Chiang, K. S., and Del Ponte, E. M. (2022a). Plant disease severity estimated visually: A century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Trop. Plant Pathol.* 47, 25–42. doi: 10.1007/s40858-021-00439-z

Bock, C. H., Pethybridge, S. J., Barbedo, J. G. A., Esker, P. D., Mahlein, A. K., and Del Ponte, E. M. (2022b). A phytopathometry glossary for the twenty-first century: Towards consistency and precision in intra- and inter-disciplinary dialogues. *Trop. Plant Pathol.* 47, 14–24. doi: 10.1007/s40858-021-00454-0

Bock, C. H., Poole, G. H., Parker, P. E., and Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Crit. Rev. Plant Sci.* 29, 59–107. doi: 10.1080/07352681003617285

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Chang, C. C., and Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 27. doi: 10.1145/1961189.1961199

Chen, X. M. (2005). Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat. *Can. J. Plant Pathol.* 27, 314–337. doi: 10.1080/07060660509507230

Chen, W. Q., Wellings, C., Chen, X. M., Kang, Z. S., and Liu, T. G. (2014). Wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici*. *Mol. Plant Pathol.* 15, 433–446. doi: 10.1111/mpp.12116

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1023/A:1022627411411

Cover, T. M., and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27. doi: 10.1109/TIT.1967.1053964

Dong, J. Z., Zeng, S. M., and Ma, Q. X. (1990). The relationships between incidence and severity of leaf rust, powdery mildew and yellow rust of wheat. *Acta Phytopathol. Sin.* 20, 55–60. doi: 10.13926/j.cnki.apps.1990.01.021

Figueroa, M., Hammond-Kosack, K. E., and Solomon, P. S. (2018). A review of wheat diseases-a field perspective. *Mol. Plant Pathol.* 19, 1523–1536. doi: 10.1111/mpp.12618

Hayit, T., Erbay, H., Varçın, F., Hayit, F., and Akci, N. (2021). Determination of the severity level of yellow rust disease in wheat by using convolutional neural networks. *J. Plant Pathol.* 103, 923–934. doi: 10.1007/s42161-021-00886-2

Huang, M. Y., Huang, W. J., Liu, L. Y., Huang, Y. D., Wang, J. H., Zhao, C. J., et al. (2004). Spectral reflectance feature of winter wheat single leaf infected with stripe rust and severity level inversion. *Trans. Chin. Soc Agric. Eng.* 20, 176–180. doi: 10.3321/j.issn:1002-6819.2004.01.042

Jiang, X. M., Feng, H. K., Chang, H., Yang, G. J., and Yang, X. D. (2021). Classification method of wheat stripe rust disease degree based on digital image. *Jiangsu Agric. Sci.* 49, 109–115. doi: 10.15889/j.issn.1002-1302.2021.23.020

Jiang, Q., Wang, H. L., and Wang, H. G. (2022). Two new methods for severity assessment of wheat stripe rust caused by *Puccinia striiformis* f. sp. *tritici*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1002627

Lamari, L. (2008). *Assess 2.0: Image analysis software for disease quantification* (Saint Paul, MN, USA: The American Phytopathological Society).

Li, G. L., Ma, Z. H., and Wang, H. G. (2011). An automatic grading method of severity of single leaf infected with grape downy mildew based on image processing. *J. China Agric. Univ.* 16, 88–93. doi: 10.11841/j.issn.1007-4333.2011.06.014

Li, X. L., Qin, F., Zhao, L. L., Li, J. H., Ma, Z. H., and Wang, H. G. (2015). Identification and classification of disease severity of wheat stripe rust using near infrared spectroscopy technology. *Spectrosc. Spect. Anal.* 35 (2), 367–371. doi: 10.3964/j.issn.1000-0593(2015)02-0367-05

Li, Z. Q., and Zeng, S. M. (2002). *Wheat rusts in China* (Beijing, China: China Agriculture Press).

Line, R. F. (2002). Stripe rust of wheat and barley in north America: A retrospective historical review. *Annu. Rev. Phytopathol.* 40, 75–118. doi: 10.1146/annurev.phyto.40.020102.111645

Liu, W. C., Wang, B. T., Zhao, Z. H., Li, Y., and Kang, Z. S. (2022). Historical review and countermeasures of wheat stripe rust epidemics in China. *China Plant Prot.* 42, 21–27, 41. doi: 10.3969/j.issn.1672-6820.2022.06.005

MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations," in *Proceedings of 5th Berkeley symposium on mathematical statistics and probability* (Los Angeles, CA, USA: University of California Press), 281–297.

Mahlein, A. K. (2016). Plant disease detection by imaging sensors - parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251. doi: 10.1094/PDIS-03-15-0340-FE

Nutter, F. W.Jr., Teng, P. S., and Shokes, F. M. (1991). Disease assessment terms and concepts. *Plant Dis.* 75, 1187–1188.

Olivoto, T., Andrade, S. M. P., and Del Ponte, E. M. (2022). Measuring plant disease severity in r: Introducing and evaluating the pliman package. *Trop. Plant Pathol.* 47, 95–104. doi: 10.1007/s40858-021-00487-5

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Pethybridge, S. J., and Nelson, S. C. (2015). Leaf doctor: A new portable application for quantifying plant disease severity. *Plant Dis.* 99, 1310–1316. doi: 10.1094/PDIS-03-15-0319-RE

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Shang, H. S., Ren, W. L., and Wang, S. M. (1990). A standard grading for wheat stripe rust disease. *Plant Prot.* 16, 31.

Shi, J. B., and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 888–905. doi: 10.1109/34.868688

von Luxburg, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* 17, 395–416. doi: 10.1007/s11222-007-9033-z

Wang, H. G. (2022). Smart phytoprotection and suggestions for its development. *J. China Agric. Univ.* 27, 1–21. doi: 10.11841/j.issn.1007-4333.2022.10.01

Wang, X. J., Ma, Z. H., Jiang, Y. Y., Shi, S. D., Liu, W. C., Zeng, J., et al. (2014). Modeling of the overwintering distribution of *Puccinia striiformis* f. sp. tritici based on meteorological data from 2001 to 2012 in China. *Front. Agr. Sci. Eng.* 1, 223–235. doi: 10.15302/J-FASE-2014025

Wang, H. G., Ma, Z. H., Wang, T., Cai, C. J., An, H., and Zhang, L. D. (2007). Application of hyperspectral data to the classification and identification of severity of wheat stripe rust. *Spectrosc. Spect. Anal.* 27, 1811–1814. doi: 10.3964/j.issn.1000-0593.2007.09.037

Wang, H., Qin, F., Ruan, L., Wang, R., Liu, Q., Ma, Z. H., et al. (2016). Identification and severity determination of wheat stripe rust and wheat leaf rust based on hyperspectral data acquired using a black-paper-based measuring method. *PloS One* 11, e0154648. doi: 10.1371/journal.pone.0154648

Wellings, C. R. (2011). Global status of stripe rust: A review of historical and current threats. *Euphytica* 179, 129–141. doi: 10.1007/s10681-011-0360-y

Zhao, J. L., Huang, L. S., Huang, W. J., Zhang, D. Y., Yuan, L., Zhang, J. C., et al. (2014). Hyperspectral measurements of severity of stripe rust on individual wheat leaves. *Eur. J. Plant Pathol.* 139, 407–417. doi: 10.1007/s10658-014-0397-6

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 20 March 2023
DOI 10.3389/fpls.2023.1135718

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Zhaoyu Zhai,
Nanjing Agricultural University, China
Jun Yan,
China Agricultural University, China

*CORRESPONDENCE
Lei Feng
Lfeng@zju.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 01 January 2023
ACCEPTED 21 February 2023
PUBLISHED 20 March 2023

CITATION
Tang W, Wu N, Xiao Q, Chen S, Gao P, He Y and Feng L (2023) Early detection of cotton verticillium wilt based on root magnetic resonance images. *Front. Plant Sci.* 14:1135718. doi: 10.3389/fpls.2023.1135718

COPYRIGHT
© 2023 Tang, Wu, Xiao, Chen, Gao, He and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Early detection of cotton verticillium wilt based on root magnetic resonance images

Wentan Tang[1,2†], Na Wu[3†], Qinlin Xiao[1,2], Sishi Chen[1,2], Pan Gao[4], Yong He[1,2] and Lei Feng[1,2*]

[1]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2]Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou, China, [3]School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China, [4]College of Information Science and Technology, Shihezi University, Shihezi, China

Verticillium wilt (VW) is often referred to as the cancer of cotton and it has a detrimental effect on cotton yield and quality. Since the root system is the first to be infested, it is feasible to detect VW by root analysis in the early stages of the disease. In recent years, with the update of computing equipment and the emergence of large-scale high-quality data sets, deep learning has achieved remarkable results in computer vision tasks. However, in some specific areas, such as cotton root MRI image task processing, it will bring some challenges. For example, the data imbalance problem (there is a serious imbalance between the cotton root and the background in the segmentation task) makes it difficult for existing algorithms to segment the target. In this paper, we proposed two new methods to solve these problems. The effectiveness of the algorithms was verified by experimental results. The results showed that the new segmentation model improved the Dice and mIoU by 46% and 44% compared with the original model. And this model could segment MRI images of rapeseed root cross-sections well with good robustness and scalability. The new classification model improved the accuracy by 34.9% over the original model. The recall score and F1 score increased by 59% and 42%, respectively. The results of this paper indicate that MRI and deep learning have the potential for non-destructive early detection of VW diseases in cotton.

KEYWORDS

cotton, verticillium wilt, root, MRI, detection, SwinUNet, ResNet

1 Introduction

Cotton is an essential cash crop. Unfortunately, cotton's growth can be affected by numerous diseases, with Verticillium wilt (VW) being the most destructive (Billah et al., 2021). VW is a systemic disease of the entire reproductive period, with symptoms typically appearing after bud emergence and peaking during flowering and boll set (Li et al., 2021). Verticillium dahliae (Vd), a soil-borne fungus with a wide range of hosts and high pathogenicity, is the primary cause of VW disease in cotton regions (Shaban et al., 2018). It

is worth noting that the Vd primarily infects cotton plants from the root systems upward. Therefore, the detection of VW is of significant importance for reducing the devastation and economic loss.

The traditional method, such as the polymerase chain reaction procedure (Altaae, 2019), for detecting VW disease in cotton is the chemical detection method. These methods are destructive and relatively time-consuming. With the advancement of technology, some non-destructive detection techniques have been brought up, such as hyperspectral imaging and thermal imaging (Poblete et al., 2021; Yang et al., 2022). However, they can only detect VW based on symptoms in the above-ground parts of the plant. Since VW is infested from the roots, detection from the roots can be better for early detection.

Researchers have proposed various non-destructive methods to study the root systems, such as hydroponics, water-cooled gel culture, and computed tomography (CT) imaging. CT can detect roots *in situ* in soils such as wheat (Gregory et al., 2003), corn (Lontoc-Roy et al., 2006), and rice (Rogers et al., 2016). However, the similar absorption coefficients of soil and roots made it difficult for CT to distinguish them (Wu et al., 2018). In recent years, Magnetic Resonance Imaging (MRI) has been applied to the non-destructive inspection of plant roots. The principle of MRI is to obtain information by acquiring magnetic resonance signals at various locations within a magnetic field and then reconstructing the image of the object's interior. The technique is extremely effective at detecting hydrogen atoms within a substance. During imaging, the signal intensity of spatial voxels is proportional to the number of hydrogen atoms present in the sample (Kumar Patel et al., 2015; Li et al., 2018; Lu et al., 2019). Both CT and MRI could produce tomographic images, but the results of experiments indicated that MRI provides greater root systems detail (Metzner et al., 2015). MRI has a high resolution, a variety of imaging parameters, the ability to choose any angle and dimension, and no radiation damage to the sample. Compare to medical MRI instruments, low-field nuclear magnetic resonance (LF- NMR) used in this paper is much cheaper. LF- NMR instrument has been widely used in agricultural science, including wheat(Chao et al., 2020), rice (Song et al., 2021), maize (Song et al., 2022), etc. In addition, the imaging parameters of medical MRI instruments are fixed parameters pre-set to obtain images of the inside of the body, while the parameters of LF-MRI can be flexibly adjusted. Therefore, LF-MRI technology is available for early, *in situ* detection of plant root diseases.

Since the morphology of plant roots changes after being affected by pathogens and external stresses, the morphological characteristics of roots in MRI images can be used to detect plant root diseases. Si mone Schmittgen et al. found by MRI that the volumetric growth of the taproot had already started to decrease on the fourteenth day after foliar Cercospora inoculation (Schmittgen et al., 2015). C.Hillnhütter et al. used MRI to non-invasively detect subsurface symptoms of sugar beet crown and root rot caused by sugar beet cyst nematodes and rhizobia. Lateral root development and sugar beet deformation were evident on MRI images of beet cyst nematode-infected plants 28 days after inoculation compared to uninfected plants (Hillnhuetter et al., 2012). Nowadays, some scholars have used deep learning and transfer learning to segment the plant root system and detect plant disease based on leaf image data. With a public dataset of 54,306 diseased and healthy plant leaves that were collected under controlled conditions, Sharada P.Mohanty et al. trained deep convolutional neural networks (CNN) and employed transfer learning to identify 14 crops and 26 diseases (or lack thereof) (Mohanty et al., 2016). Both (Wang et al., 2022) and (Guo et al., 2022) works of literature improved Swin Transformer (SwinT) to achieve the detection of plant diseases with an accuracy of 98.97% and 98.2%, respectively.

Compared with existing models, there were two difficulties in this paper. First, this paper studied the transverse section of the root system, which was different from the features of the longitudinal pictures of the root system in previous works. These segmentation models could not directly extract the features of root system cross-section in MRI images well. And the ratio of pixels occupied by the root system and soil studied in this paper was too disparate, which made the existing advanced segmentation models only segment the soil correctly and unable to capture the features of the root system. Second, existing disease detection models were mainly for RGB images of leaves and stems. However, in this paper, MRI images were grayscale images, which had less information than RGB images. Moreover, the disease features of leaves and stems were more numerous and obvious than those of roots. The MRI images of cotton roots could not provide so many features information on which the existing classification models were based. And the number of images of healthy and diseased samples is different, which can lead to a large loss in the model training process.

In this paper, the main purpose was to investigate the feasibility of MRI-based detection of VW infestation from cotton roots system. The specific objectives included the following: (1) denoise MRI images of cotton root to improve the signal-to-noise ratio of the images; (2) modify the MRI images segmentation model for obtaining the root target; (3) improve the image classification model to classify root MRI images between healthy and infected by Vd.

The main contributions include the following:

- The influence of pre-processing methods of cotton root MRI images was compared.
- We proposed the segmentation model and early disease detection model applicable to the MRI images of cotton roots. These models addressed the problems of unbalanced soil and root pixel scales and small data sets.
- Compared with other advanced models, our new models showed better robustness and extensibility. This demonstrated that early detection of cotton VW based on cotton root MRI images and deep learning was feasible.

The structure of the remaining portion of this paper is as follows: Section 2 describes the materials and methods. Section 3 explains the results and provides a discussion, and finally, conclusions are given in Section 4.

2 Materials and methods

2.1 Sample preparation

In May 2022, the experiment was conducted at the college of Biosystems Engineering and Food Science at Zhejiang University in

Hangzhou, Zhejiang province, China. The cultivar of cotton and oilseed rape were tested: Xinluzao 45 and Zhongshuang11, respectively. Cotton and oilseed rape seeds and the conidia solution of Vd were provided by the Agricultural College of Shihezi University, China. The concentration of conidia of Vd in the solution was 10^6 conidia per ml. The cotton was divided into two groups: an experimental group and a control group of 20 plants each. These two groups received identical quantities of watering and fertilization. Each cotton plant in the experimental group was injected with 40 ml of a conidia solution. The control group was replaced with an equal amount of sterile water. After inoculation, the cotton was transferred to a greenhouse with daytime temperatures of 26°C and nighttime temperatures of 24°C and 60% humidity. MRI images of the root systems of 10 healthy and 10 infected cotton plants were collected on both day 15 and day 45 after inoculation. To avoid the effect of high soil moisture content on MRI imaging, the cotton was not watered for 48 hours before the formal MRI experiment. If the soil has high water content, it will be difficult to distinguish between the soil and tiny lateral roots. As shown in Figure 1B, a low-field magnetic resonance instrument (MesoMR23-060V-I, Niumag Co., Ltd., Suzhou, China) was utilized to acquire MRI images of cotton root systems. The low-field MRI device cannot collect images of targets smaller than 1 mm. The instrument relies primarily on the moisture signal for imaging. The more moisture a sample contains, the brighter it appears in the MRI images.

2.2 MRI images acquisition and data set division

Image acquisition: As shown in Figure 1A the entire cotton plant, including the soil, was placed in a 60 mm sample tube made of temperature-resistant quartz material. Spin-echo (SE) sequences were used to acquire axial MRI images. To obtain higher-quality MRI images, the following imaging parameters of the SE were optimized based on imaging quality and imaging time: TR (Repetition Time) = 1100 ms, TE (Echo Time) = 18.14 ms, Averages (Accumulation times at pre-scan) = 4, Slice thickness = 2 mm, Slice gap = 0.5 mm. The 2D Fourier transform reconstruction method built into the imaging software is used to reconstruct the image, after which 256×256 grayscale images were saved. 1191 MRI images were obtained, including 635 images of healthy roots and 556 images of infected roots. Samples were also collected from 2 healthy rape roots that had been growing for about 20 days, with a total of 32 MRI images.

Data set division: The dataset was divided based on the proportion 8:2 in this paper. In the image segmentation task, five healthy cotton root systems and five root systems infected with Vd were randomly selected. The 315 images of these ten cotton root systems were collected and utilized as a dataset for the segmentation task after being denoised. 252 images were used as the training set and 63 images were treated as the testing set. In the image



FIGURE 1
(A) Cotton experimental samples; **(B)** The low-field MRI instrument used in this paper.

classification task, 1191 images were used as the dataset, the training set consisted of 953 images and the testing set contained 238 images.

2.3 Data analysis

2.3.1 Fine-tuning

Existing models in supervised learning require large quantities of labeled data, computational time, and resources. To save time and effort, transfer learning for deep learning is gaining more and more attention (Jiang et al., 2022). Transfer learning aims to apply knowledge or patterns acquired in one domain or task to a distinct but related domain or problem. Fine-tuning model is a method of transfer learning. The model parameters of pre-trained models are superior to those obtained by others after training with some classic models (VGG16/19, ResNet) and utilizing large datasets as training sets (ImageNet, COCO) (Hasan et al., 2022). In this experiment, both MRSwinUNet and MRResNet models utilized the fine-tuning method. After retaining the architecture of the model, the model was retrained using the initial weights of the pre-trained model to fine-tune.

2.3.2 Loss function

The Focal loss (Lin et al., 2017) is a loss function that deals with the imbalance of sample classification. It focuses on adding weights to the losses corresponding to the samples according to the ease of sample discrimination, i.e., adding smaller weights to the samples that are easy to distinguish and larger weights to the samples that are difficult to differ. The Focal loss function was improved from the cross-entropy loss function. As in Equation (1)

$$\mathbf{CE}(\mathbf{p},\mathbf{y}) = \begin{cases} -\log(p), & if \;\; y = 1 \\ -\log(1-p), & otherwise \end{cases} \tag{1}$$

Here, $\boldsymbol{y}$ takes values of 1 and -1, representing the foreground and background, respectively. $\boldsymbol{p}$ takes values ranging from 0 to 1 and is the probability that the model predicts belonging to the foreground.

Next, as shown in Equation (2), a function on $\boldsymbol{p}$ is defined.

$$p_t = \begin{cases} p, & if \;\; y = 1 \\ 1-p, & otherwise \end{cases} \tag{2}$$

The combination of equation (1) and equation (2) leads to the simplified equation (3).

$$\mathbf{CE}(\mathbf{p},\mathbf{y}) = \mathbf{CE}(p_t) = -\mathbf{log}(p_t) \tag{3}$$

To solve the positive and negative sample imbalance problem, a weighting factor $\boldsymbol{\alpha}$ is introduced belonging to [0,1]. When it is a positive sample, the weighting factor is $\boldsymbol{\alpha}$, and when it is a negative sample, the weighting factor is 1-$\boldsymbol{\alpha}$. The loss function can be rewritten as:

$$\mathbf{CE}(p_t) = -\alpha_t \mathbf{log}(p_t) \tag{4}$$

Formula (4) is called balanced cross entropy(BCE) loss and is the baseline for proposing Focal loss.

BCE loss does not distinguish between simple or difficult samples. When the number of easy-to-distinguish negative samples is super high, the whole training process will revolve around the easy-to-distinguish negative samples, which will in turn swamp the positive samples and cause large losses. Therefore, a modulation factor is introduced here to focus on the hard-to-score samples with the following formula (5).

$$\mathbf{FL}(p_t) = -(1-p_t)^{\gamma} \mathbf{log}(p_t) \tag{5}$$

$\boldsymbol{\gamma}$ is a parameter in the range [0, 5]. $(1\text{-}\boldsymbol{p_t})^{\boldsymbol{\gamma}}$ can reduce the loss contribution of the easy-to-score samples and increase the loss proportion of the hard-to-score samples. When $\boldsymbol{p_t}$ tends to 1, which means that the sample is easily distinguishable. Then the modulating factor $(1\text{-}\boldsymbol{p_t})^{\boldsymbol{\gamma}}$ tends to 0, which means that it contributes less to the loss, i.e., it reduces the proportion of loss of the easily distinguishable sample. Small $\boldsymbol{p_t}$ means that if a sample is divided into positive samples, but the probability that the sample is positive is particularly small, the modulating factor $(1\text{-}\boldsymbol{p_t})^{\boldsymbol{\gamma}}$ tends to 1, which does not have much effect on the Loss.

By balancing the above for positive and negative samples as well as difficult and easy samples, the final Focal loss formula (6) should be obtained.

$$\mathbf{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma} \mathbf{log}(p_t) \tag{8}$$

The imbalance in the number of positive and negative samples can be suppressed by $\boldsymbol{\alpha_t}$. And the imbalance in the number of simple or difficult-to-distinguish samples can be controlled by $\boldsymbol{\gamma}$. In this experiment, $\boldsymbol{\gamma}$ is 2 and $\boldsymbol{\alpha_t}$ is 0.25.

2.3.3 Segmentation models

A hierarchical transformer called SwinT has been proposed (Liu et al., 2021), which was based on shift windows to implement the computation. The move operation allowed adjacent windows to be interacted with, significantly reducing the computational complexity. Compared with CNN, it showed competitive or even better performance on various visual benchmarks.

SwinUnet (Cao et al., 2021) was based on the SwinT network design for the image segmentation task, having transformer modules similar to the UNet structure. Supplementary Figure 1A represents the structure diagram of SwinT for the classification of the ImageNet dataset. And Supplementary Figure 1C depicts two SwinT modules connected in series, like a traditional multi-headed self-attentive (MSA) module structure's construction on shifted windows. Each SwinT module includes a layer normalization layer (LN), a MSA module, a residual connection, and a multilayer perceptron with an activation function. In two consecutive transformer modules, a window-based multi-headed self-attentive (W-MSA) module and a shifted-window-based multi-headed self-attentive module (SW-MSA) are applied, respectively. $\hat{z}^l$ and z^l denote the ML of the (SW-MSA) module and the 1 th block, respectively. The number of operations required to compute the correlation between two locations did not increase with distance, which made it possible to capture global semantic information more efficiently. In this paper, three major improvements were made to the SwinUNet model to obtain the model named MRSwinUNet.

First, We used transfer learning to better train the model. The specific step of the fine-tuning technique for transfer learning was to first preserve the original structure and then train with pre-trained weights. This improvement saved time on label annotation and reduced the requirement for the number of datasets. Next, the BCE loss function was replaced by the Focal loss function, which could distinguish the difficulty of segmented samples. A higher weight was given to the more difficult segmented roots, while a lower weight was given to the easily segmented soil pixels. The problem of large differences in the proportion of pixels occupied by cotton roots and background soil was solved. Finally, the AdamW optimizer was applied to improve the performance of the network.

2.3.4 Classification models

The residual block structure of the ResNet network was proposed to solve the problem of gradient disappearance or gradient explosion. At the same time, it also addressed the issue of deeper levels leading to network performance degradation. Therefore, ResNet34 (He et al., 2016) network was chosen as the classification model based on the size of our dataset and the network performance of the devices used. Supplementary Figure 2C is a specific presentation of the residual block in Supplementary Figure 2A. In Supplementary Figure 2C, the feature matrix obtained after a series of convolutional layers on the mainline is summed with the input feature matrix, which is then output by the activation function. The output feature matrix shape of the main branch and shortcut must be the same. Formally, the desired underlying mapping is denoted as $\mathbf{H}(\mathbf{x})$ and the stacked nonlinear layers are made to fit another mapping: $\mathbf{F}(\mathbf{x}) = \mathbf{H}(\mathbf{x}) - \mathbf{x}$. The original mapping is reshaped as $\mathbf{F}(\mathbf{x}) + \mathbf{x}$. It is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. In this work, we made three improvements based on the ResNet model using the cotton root MRI image dataset. The model named MRResNet was obtained afterwards.

We used transfer learning and changed the loss function of ResNet to Focal loss function, and replaced the original optimizer with AdaBound to solve the problem that the MRI images of cotton roots have less information than the RGB images of leaves or stems. The issue of different number of MRI images for healthy and diseased samples was also addressed.

The training parameters for the segmentation and classification network models are shown in Table 1.

2.4 Model evaluation and software

This paper evaluated the denoising model using the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) indices. Given a clean image I and a noisy image K of size m×n, the mean square error (MSE) and PSNR is defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{4}$$

$$PSNR = 10 \times log_{10}\left(\frac{MAX_I^2}{MSE + \varepsilon}\right) \tag{5}$$

where MAX_I^2 is 255. ε is a very small constant that prevents the denominator from being zero. SSIM indicates the degree of similarity between two images. The definition is as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{6}$$

where x and y are two signal indicators, μ_x and μ_y represent the means of x and y respectively, and σ_x and σ_y represent the standard deviations of x and y, respectively. σ_{xy} represents the covariance of x and y. And c_1,c_2,c_3 are constants to avoid systematic errors brought by a zero denominator.

In the segmentation task, the Dice coefficient (1), mean Intersection over Union (mIoU), Recall, and Precision metrics were used. And Accuracy, F1 score, Recall, and Precision metrics were used in the classification task.

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{TP + FP + FN} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$F1\ \ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

TP, ***TN***, ***FP***, and ***FN*** indicate the number of true positives, true negatives, false positives, and false negatives, respectively. ***k*** is the total number of categories to be segmented.

TABLE 1 The segmentation and classification network model training parameters.

Items	Segmentation values		Classification values	
	original model	MRSwinUNet	original model	MRResNet
Learning rate	0.0001	0.0001	0.0001	0.0001
Optimizer	SGDM	AdamW	SGDM	AdaBound
Num_workers	4	4	4	4
Loss function	BCE loss	Focal loss	BCE loss	Focal loss
Epochs	100	100	10	10

The cotton root systems *in situ* images were annotated by the lasso tool of Adobe Photoshop CC2020 in the segmentation task. The segmentation and classification models were developed using the deep learning framework PyTorch (version 1.7.1). All models were generated with PyCharm (version 2019.2.3). A custom-built workstation with 48 GB of RAM and two GTX 1080 Ti graphics cards (NVIDIA, California, United States) was utilized.

3 Results and discussion

3.1 Denoising of the cotton root's MRI images

Image denoising could reduce the damage of noise to make the root system features clearer. In this paper, the parameters were optimized and the best parameter results were obtained for three models NLM, GLPF, and MF. Table 2 details the comparison of the effects of each model.

It is well known that a higher PSNR value represents a cleaner image. SSIM ranges from 0 to 1, with values closer to 1 indicating more image detail retention. NLM had the highest PSNR score and SSIM with 32.776 dB and 0.952, respectively. The PSNR score of GLPF and MF did not exceed 30 dB. Meanwhile, the SSIM index of GLPF and MF did not exceed 0.9.

To examine the effect of denoising each model more visually, Figure 2 presents the sample images for each model. NLM appeared the least noisy, with a clear background and more complete details. The denoising of GLPF and MF blurred the image and left the details incomplete.

Based on the PSNR and SSIM metrics and the subjective judgment of the vision, NLM had the best denoising effect and the most detail retention. It was because it had the ability to calculate the required pixels by weighted averaging of the entire pixels of the image, thus reducing the loss of image details. Since the noise was primarily concentrated in the high-frequency band, GLPF filtered the noise information to make the image smooth. But it also blurred the image. Additionally, MF also made the image more blurred. In this paper, it was considered that the image blurring and detail loss caused by GLPF and MF denoising processes were unacceptable. Therefore, the NLM model was chosen to denoise the MRI images.

3.2 Segmentation of the cotton root's MRI images

In the segmentation task, the clean images after denoising were segmented to extract the root systems region, which was beneficial for the subsequent classification of the root systems. And the images were segmented at the pixel level.

Table 3 outlines the segmentation results. The Dice coefficients and mIoU of both SwinUNet-B and TransUNet-B models were 0.5, and their precision and recall were both 0. It indicates that SwinUNet-B and TransUNet-B are not directly applicable to the segmentation task of cotton root MRI images. After improving the model, MRSwinUNet and TransUNet-D performed well with all metrics close to each other. However, the training time of MRSwinUNet was longer than the MRSwinUNet. This was a big drawback of the TransUNet model. So our MRSwinUNet model had the best overall performance. Compared with the original model SwinUNet-B, the Dice coefficient and mIoU of out model increased by 46% and 44%, respectively.

To further demonstrate the results in Table 3, Precision-Recall (PR) curves were given in this paper. In the PR curve image, the closer the curve is to the coordinate (1,1), the better the performance. In Figure 3, TransUNet had the best PR curve, but MRSwinUNet's PR curve was right next to it and got just as good results. And the original model SwinUNet performed the worst.

According to the observed experimental images, it was known that the ratio of pixels occupied by the root system and the soil was approximated at a minimum of 1:16383. However, the original SwinUNet-B and TransUNet-B models were trained by assigning the same weights to the root system and the soil. In this case, the original loss function and the optimizer only guided the model to correctly segment the soil pixels and could not work for the root system roots. In addition, although the metrics of both TransUNet-D and SwinUNet-D were improved, they were still not as good as the combined performance of MRSwinUNet. It was probably due to the reason that the Dice loss function and Adam optimizer did not perform as well as the Focal loss function and AdamW used in this paper.

To better demonstrate the segmentation effect, we performed a visual evaluation. Figure 4 presents the representative figure of root segmentation effect of SwinUNet-D, TransUNet-D, and MRSwinUNet. The findings demonstrate that all three sample maps differed somewhat from the accurate label maps in detail. For instance, inside

TABLE 2 The denoising results of the MRI images of the cotton root systems.

Index	Optimal results	PSNR(dB)	SSIM
NLM	h = 10 Search windowsize = 21 x 21 Patch window size=3	32.776	0.952
GLPF	Kernel size = 5 x 5	24.643	0.882
MF	Kernel size =7 x 7	22.650	0.807

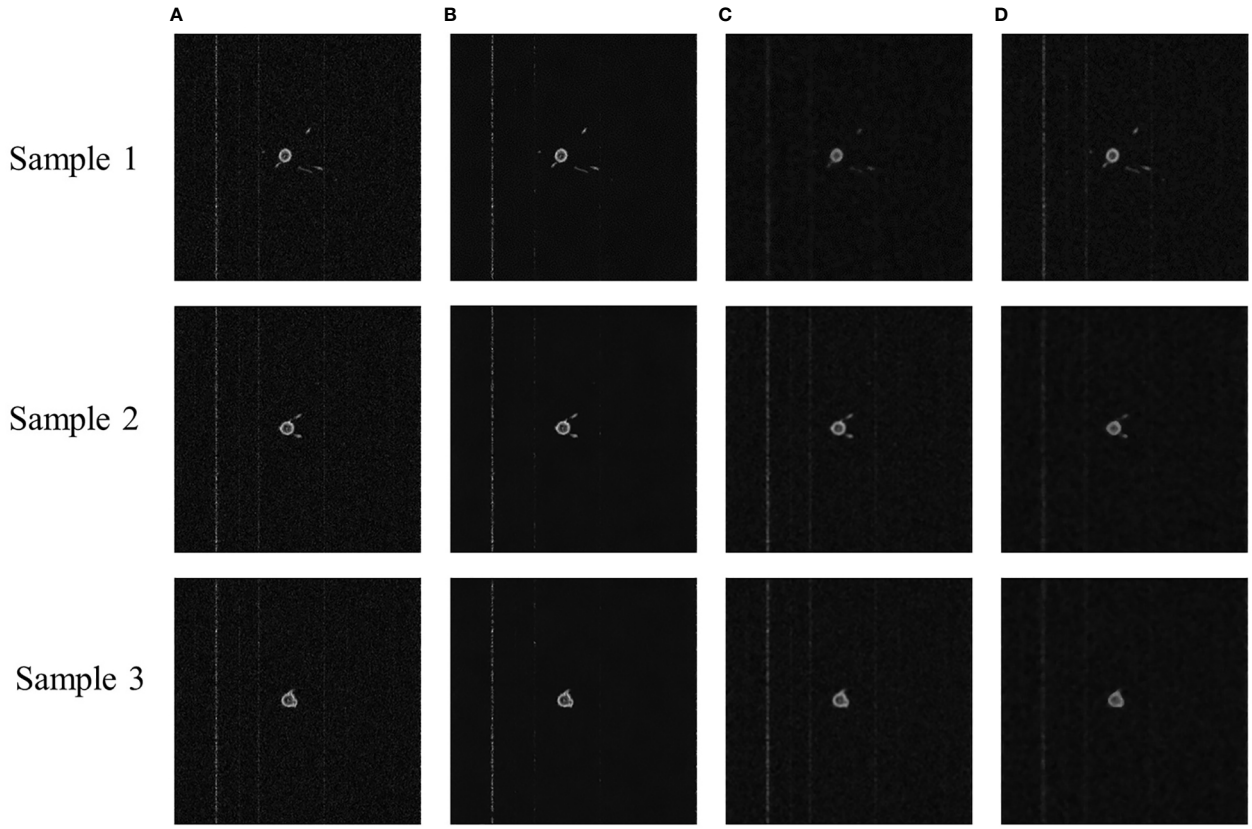


FIGURE 2
Sample images of denoised MRI images of cotton root system. **(A)** Original MRI image, **(B)** NLM denoised image, **(C)** GLPF denoised image, and **(D)** MF denoised image.

TABLE 3 The segmentation results of the MRI images of the cotton root systems.

model	Dice	mIoU	Precision	Recall	Time(s)
TransUNet - B	0.50	0.50	0	0	3360
TransUNet - D	0.96	0.93	0.95	0.90	3624
SwinUNet - B	0.50	0.50	0	0	1685
SwinUNet - D	0.95	0.90	0.89	0.90	1683
MRSwinUNet	0.96	0.94	0.91	0.95	1659
MRSwinUNet- Canola	0.93	0.90	0.95	0.89	-

The symbol "-" means no calculation time.

the red box of Figure 4, the SwinUNet-D barely segmented any effective information on the root. The resultant map of MRSwinUNet segmentation was similar to the original label with sufficient detail. So MRSwinUNet was considered to be the most optimal model for the overall performance of the segmentation task.

To investigate the scalability of the MRSwinUNet model, we selected 32 MRI images of canola obtained with the same acquisition method and preprocessing method. The trained MRSwinUNet model was used to segment the rape dataset, and the results were displayed in the MRSwinUNet-Canola model in Table 3, the Dice, mIoU, of the canola segmentation results were 0.93 and 0.90, respectively. In addition, its Precision score was 4% higher than that of the cotton dataset. This indicates that MRSwinUNet has better robustness and extensibility.

Previous research scholars (Shen et al., 2020; Kang et al., 2021; Lu et al., 2022; Zhao et al., 2022) have also done comprehensive studies on plant root segmentation. In this paper (Kang et al., 2021), the cotton mature root systems were used as the research object. They designed a semantic segmentation model of cotton roots *in-situ* images based on the attention mechanism. The precision and recall values were 8.7% and 4.8% higher than those in this paper, respectively. This would be due to the high resolution (10200×14039 dpi) of the root images they acquired, which was easy to identify and segment. In addition, they trained the model directly using their dataset. Although the training

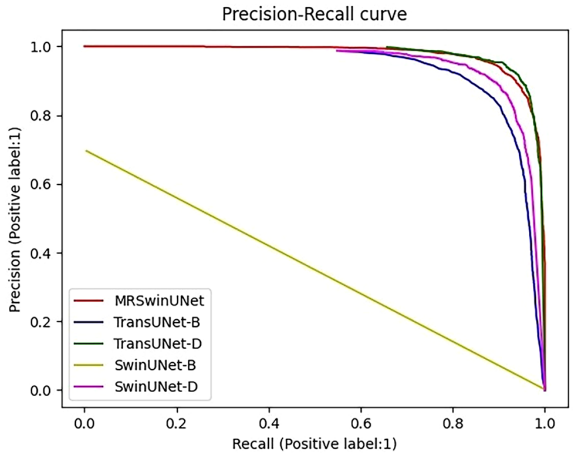


FIGURE 3
Precision-Recall curve. The label is 1, which means the infested root system is the positive sample. MRSwinUNet is our improved segmentation model. The TransUNet-B and TransUNet-D represent the TransUNet models using BCE loss and SGDM, Dice loss, and Adam, respectively. SwinUNet-B and SwinUNet-D represent the SwinUNet models using BCE loss and SGDM, Dice loss, and Adam, respectively.

process took a lot of time, it facilitated the extraction of root features in the images and reduced segmentation errors.

3.3 Classification of the cotton root's MRI images

Figures 5, 6 show the MRI images of the healthy and infected root systems. The smallest root diameter that can be detected by the low-field MRI instrument used is 1 mm. This means that all the roots in the figure have a cross-sectional diameter greater than or equal to 1 mm. It should be noted that the healthy root systems had more branching cross-sections than the diseased root systems, which could be ascribed to the fungus also colonizing the ducts and secreting toxins that damage the cells (Bai et al., 2022; Lv et al., 2022; Ren et al., 2022; Sayari et al., 2022). Consequently, cell growth would be hindered, and the number of lateral roots reduces, which provides the possibility of classifying the MRI images of healthy and unhealthy root systems.

In Table 4, the results of all metrics of the original models (SwinT, Vgg16Net, ResNet) were unsatisfactory. It indicates that the original models cannot perform the classification task regarding the root MR images. Compared to the original model, the results obtained by our MRResNet using all five preprocessing methods were significantly improved. The highest accuracy was achieved when MRResNet used the dataset processed by denoising first and then segmenting, with 34.9% improvement over the original ResNet model, and 59% and 42% improvement for Recall and F1, respectively. When MRResNet was trained on the dataset processed in the other four ways, the results were all improved over the original model. But it was still lower than the results of the dataset processed by the denoising-only method, denoising first and then the segmentation method.

To compare more comprehensively the effect of image preprocessing methods on the classification results of MRResNet models, PR curves were plotted. In Figure 7, the curve of the denoised and then-segmented dataset was closest to the coordinate (1, 1). This indicates that this dataset performs best in the classification task. From the results, it can be concluded that the denoised and then-segmented dataset worked best in classification model training. Because it filtered out the noise, reduced image

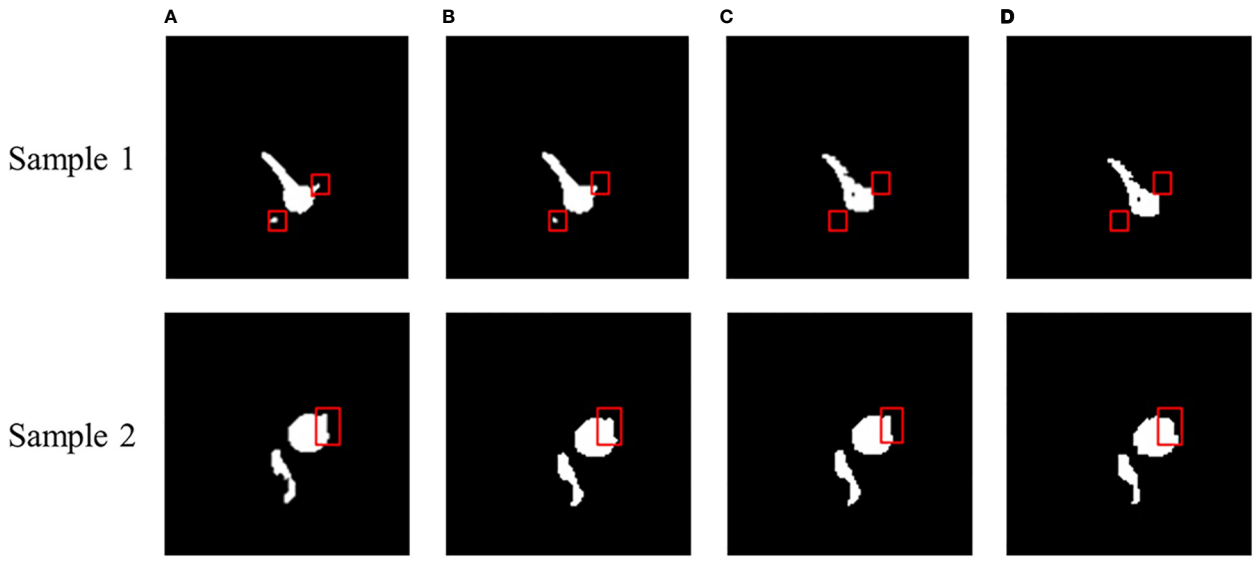


FIGURE 4
Sample segmentation effect of MRI images. **(A)** Real label, **(B)** MRSwinUNet, **(C)** TransUNet-D, **(D)** SwinUNet-D.

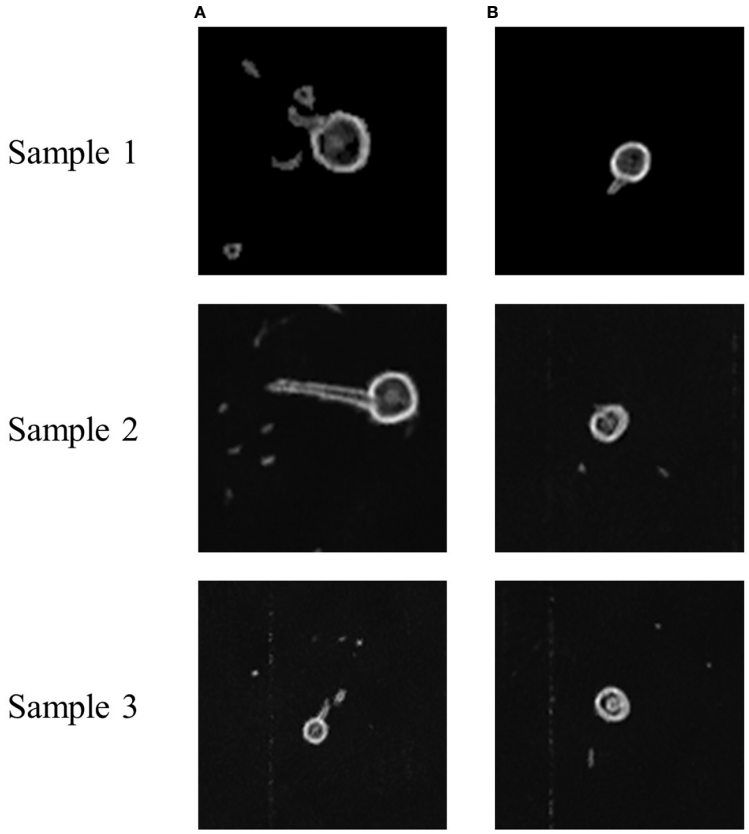


FIGURE 5
Sample MRI images of healthy and diseased roots of cotton. **(A)** Healthy cotton root system, **(B)** Infected cotton root systems. These sample images are from the same location of different root systems.

pollution, and avoided the problem of blurred root features. Furthermore, the root targets were extracted precisely by segmentation, which made the root features more clearly. The denoised dataset performed the second best. This was because the image denoising process mainly filtered out the noise in the image, but some root features had weak signals that were not further extracted by segmentation, which caused the classification model to ignore this part of the signal. The bad thing was that the dataset with only segmentation and segmentation followed by denoising process methods lost the original root system features. The reason was that without noising processing, which made the image contaminated with noise, the segmentation model did not recognize the segmented features and lost the smaller but more important information of the signal, such as the lateral root cross-section.



FIGURE 6
Samples of the cotton root system. The root on the left is healthy. The root on the right is affected by the Vd.

TABLE 4 Classification results of the MRI images of healthy versus Vd-infested cotton root systems.

Model	No.	Accuracy (%)	Precision	Recall	F1	Time(s)
SwinT	1	53.40	0	0	0	327
	2	53.40	0	0	0	331
	3	53.40	0	0	0	318
	4	53.40	0	0	0	327
	5	53.40	0	0	0	316
Vgg16Net	1	46.60	1	1	1	357
	2	53.40	0	0	0	359
	3	53.40	0	0	0	351
	4	46.60	1	1	1	352
	5	53.40	0	0	0	355
ResNet34	1	53.60	1	0.05	0.09	246
	2	58.00	1	0.56	0.74	270
	3	53.40	0	0	0	256
	4	60.10	1	0.41	0.58	268
	5	53.40	0	0	0	255
MRResNet	1	89.90	1	1	1	259
	2	**92.00**	1	0.96	0.97	267
	3	76.50	1	0.91	0.95	252
	4	**95.00**	**1**	**1**	**1**	268
	5	70.20	1	1	1	261

The column No. represents the datasets with different preprocessing methods. Five numbers from 1 to 5 represent the original dataset, denoised dataset only, segmented dataset only, denoised-resegmented dataset, and segmented-redenoised dataset, respectively.

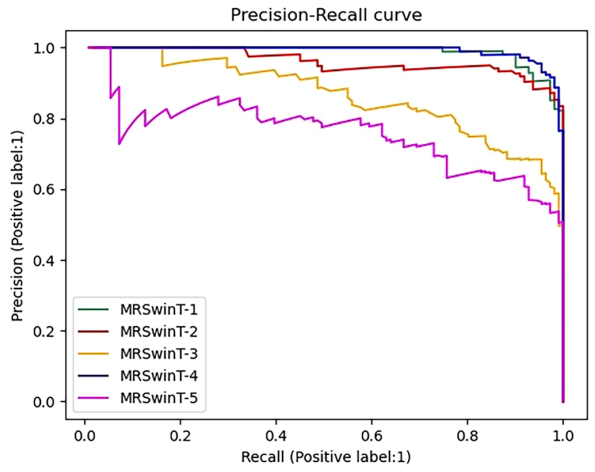


FIGURE 7
Precision-Recall curve. The label is 1, which means the infested root system is the positive sample. MRSWinT-1, MRSWinT-2, MRSWinT-3, MRSWinT-4, and MRSWinT-5 represent the original dataset, denoised dataset only, segmented dataset only, denoised-then-segmented dataset, and segmented-then-denoised dataset, respectively.

In conclusion, MRResNet was considered the optimal model considering all model metrics, training time, and PR curves. The best way to process the dataset was to denoise the images first and then segment them.

Compared with the high accuracy of the existing literature (Li et al., 2020; Liang, 2021; Santos-Rufo and Rodriguez-Jurado, 2021; Sivakumar et al., 2021; Elaraby et al., 2022; Memon et al., 2022), the accuracy of the identification of roots suffering from cotton VW disease was about 4% lower in this paper. Studies in the literature have targeted leaves and stem with obvious disease symptoms, such as leaf yellowing and wilting. Thus, the accuracy was higher when it came to disease detection. However, this paper studied cotton root systems in the early stages of VW. Since the morphology of each cotton plant varied, the classification model probably misclassified healthy cotton with a small root system as diseased cotton or, conversely, misclassified diseased cotton with a well-developed root system as healthy cotton. These misclassifications resulted in a lower accuracy rate in this paper than in other literature. Nevertheless, the method in this paper still provided a new idea for the detection of cotton VW disease. After the root system was infested, it had already changed before the leaves turned yellow and

wilted. In this situation, theoretically, the technique adopted in this paper could detect the disease much earlier.

3.4 Limitations and prospects

Image denoising and segmentation contributed to clean root systems MRI images, and deep transfer learning improved the ability to learn image features. The combination of these two approaches realized effective classify healthy and Vd-infested cotton roots. However, after being inoculated, the immune system of cotton was damaged. Along with that, there was a great possibility of infestation by other pathogens, which could be time-consuming and costly to identify. Considering the observation that cotton predominantly presented symptoms of VW when it developed, cotton VW was examined as the main disease in this paper. Besides, the number of lateral roots in this paper was only observed in 2D images. The changes in root morphology after infestation by Vd were not presented in full. In the future, we will continue to study the changes in the three-dimensional morphology of cotton roots after being infested with Vd. Finally, due to the lack of images of other plant roots affected by VW disease, there was no way to do experiments to further explore its robustness and scalability. In the future, we will collect more image data on plant roots suffering from VW disease, and thus build a robust and extensible model for the detection of VW disease.

4 Conclusions

In this paper, we first used cotton root cross-section LF-MRI images as samples to explore the feasibility of early nondestructive detection of VW disease in cotton using deep. First, the performance of three denoising models NLM, GLPF, and MF was compared, and the results showed that NLM had the best denoising effect. After that, the SwinUNet model was modified in three parts and obtained the MRSwinUNet applicable to the MRI image segmentation of the cotton root system. The Dice and mIoU of MRSwinUNet increased by 46% and 44%, respectively, over the original SwinUNet's results. And it addressed the problem of unbalanced soil and root pixel proportions and reduced the effort as in the original model. MRSwinUNet also had a good segmentation effect on MRI images of the canola root system. Subsequently, NLM and MRSwinUNet were selected to denoise and segment the cotton root dataset respectively, and the classification datasets with five pre-processing methods were obtained. And then the original classification models (SwinT, Vgg16Net, ResNet) were chosen to classify cotton root images, but the results were extremely poor. Therefore, in this paper, we made improvements to the ResNet model to obtain the MRResNet model for cotton root MRI image classification. The results of five datasets were compared on the classification model, and showed that the first denoising and then segmentation treatment worked best. When MRResNet used the best dataset, its accuracy improved by 34.9% over the original model. Meanwhile, the recall and F1 improved by 59% and 42%, respectively. This demonstrates the feasibility of detecting cotton VW disease at an early stage using deep learning and MRI images of the cotton root system. The paper provides a new research idea for the detection of VW disease in cotton.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WT, QX, and SC designed the study, conducted the experiment, and wrote the manuscript. LF and NW supervised experiments at all stages and performed revisions of the manuscript. PG and YH performed revisions of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by National Natural Science Foundation of China (31871526, 61965014) and the Shenzhen Science and Technology Projects (CJGJZD20210408092401004), and XPCC Science and Technology Projects of Key Areas (2020AB005).

Acknowledgments

The authors would like to thank Jianxun Shen (Hangzhou Raw Seed Growing Farm, Hangzhou, China) for providing the field sample cultivation site.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1135718/full#supplementary-material

References

Altaae, H. H. (2019). Using nested PCR to detect the non-defoliating pathotype of verticillium dahliae on olive orchard. *Crop Res. (Hisar)* 54 (5-6), 139–142. doi: 10.31830/2454-1761.2019.023

Bai, S., Niu, Q., Wu, Y., Xu, K., Miao, M., and Mei, J. (2022). Genome-wide identification of the NAC transcription factors in gossypium hirsutum and analysis of their responses to verticillium wilt. *Plants-Basel* 11 (19), 1–13. doi: 10.3390/plants11192661

Billah, M., Li, F., and Yang, Z. (2021). Regulatory network of cotton genes in response to salt, drought and wilt diseases (Verticillium and fusarium): Progress and perspective. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.759245

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., et al. (2021). Swin-unet: Unet-like pure transformer for medical image segmentation. *ECCV Workshops*. 205–218. doi: 10.1007/978-3-031-25066-8_9

Chao, J., Li, W., Shaowu, Y., Chuanping, L., and Lige, T. (2020). Low-field nuclear magnetic resonance for the determination of water diffusion characteristics and activation energy of wheat drying. *Drying Technol.* 38 (7), 917–927. doi: 10.1080/07373937.2019.1599903

Elaraby, A., Hamdy, W., and Alruwaili, M. (2022). Optimization of deep learning model for plant disease detection using particle swarm optimizer. *Cmc-Computers Mater. Continua.* 71 (2), 4019–4031. doi: 10.32604/cmc.2022.022161

Gregory, P. J., Hutchison, D. J., Read, D. B., Jenneson, P. M., Gilboy, W. B., and Morton, E. J. (2003). Non-invasive imaging of roots with high resolution X-ray micro-tomography. *Plant Soil* 255 (1), 351–359. doi: 10.1023/a:1026179919689

Guo, Y., Lan, Y., and Chen, X. (2022). CST: Convolutional Swin Transformer for detecting the degree and types of plant diseases. *Comput. Electron. Agr.* 202. doi: 10.1016/j.compag.2022.107407.

Hasan, M. M., Islam, N., and Rahman, M. M. (2022). Gastrointestinal polyp detection through a fusion of contourlet transform and neural features. *J. King Saud University-Computer Inf. Sci.* 34 (3), 526–533. doi: 10.1016/j.jksuci.2019.12.013

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition", in: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778. doi: 10.1109/CVPR.2016.90

Hillnhuetter, C., Sikora, R. A., Oerke, E. C., and van Dusschoten, D. (2012). Nuclear magnetic resonance: A tool for imaging belowground damage caused by heterodera schachtii and rhizoctonia solani on sugar beet. *J. Exp. Bot.* 63 (1), 319–327. doi: 10.1093/jxb/err273

Jiang, J., Shu, Y., Wang, J., and Long, M. (2022). Transferability in deep learning: A survey. *arXiv e-prints*. abs/2201.05867. doi: 10.48550/arXiv.2201.05867

Kang, J., Liu, L., Zhang, F., Shen, C., Wang, N., and Shao, L. (2021). Semantic segmentation model of cotton roots in-situ image based on attention mechanism. *Comput. Electron. Agric.* 189, 11. doi: 10.1016/j.compag.2021.106370

Kumar Patel, K., Ali Khan, M., and Abhijit, K. (2015). Recent developments in applications of MRI techniques for foods and agricultural produce - an overview. *J. Food Sci. Technol.* 52 (1), 1–26. doi: 10.1007/s13197-012-0917-3

Li, M., Li, B., and Zhang, W. (2018). Rapid and non-invasive detection and imaging of the hydrocolloid-injected prawns with low-field NMR and MRI. *Food Chem.* 242, 16–21. doi: 10.1016/j.foodchem.2017.08.086

Li, X., Yang, C., Huang, W., Tang, J., Tian, Y., and Zhang, Q. (2020). Identification of cotton root rot by multifeature selection from sentinel-2 images using random forest. *Remote Sens.* 12 (21), 24. doi: 10.3390/rs12213504

Li, Y., Zhou, Y., Dai, P., Ren, Y., Wang, Q., and Liu, X. (2021). Cotton bsr-k1 modulates lignin deposition participating in plant resistance against verticillium dahliae and fusarium oxysporum. *Plant Growth Regul.* 95 (2), 283–292. doi: 10.1007/s10725-021-00742-4

Liang, X. (2021). Few-shot cotton leaf spots disease classification based on metric learning. *Plant Methods* 17 (1), 11. doi: 10.1186/s13007-021-00813-7

Lin, T.-Y., Goyal, P., Girshick, R. B., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*. 2980–2988. doi: 10.1109/ICCV.2017.324

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 9992–10002. doi: 10.48550/arXiv.2103.14030

Lontoc-Roy, M., Dutilleul, P., Prasher, S. O., Han, L., Brouillet, T., and Smith, D. L. (2006). Advances in the acquisition and analysis of CT scan data to isolate a crop root system from the soil medium and quantify root system complexity in 3-d space. *Geoderma* 137 (1-2), 231–241. doi: 10.1016/j.geoderma.2006.08.025

Lu, W., Song, K., Zhao, D., and Zhao, D. (2019). Advances in application of low-field nuclear magnetic resonance analysis and imaging in the field of drying of fruits and vegetables. *Packag. Food Machinery* 37 (3), 47–50. doi: 10.3969/j.issn.1005-1295.2019.03.011

Lu, W., Wang, X., and Jia, W. (2022). Root hair image processing based on deep learning and prior knowledge. *Comput. Electron. Agric.* 202, 10. doi: 10.1016/j.compag.2022.107397

Lv, J., Zhou, J., Chang, B., Zhang, Y., Feng, Z., Wei, F., et al. (2022). Two metalloproteases VdM35-1 and VdASPF2 from verticillium dahliae are required for fungal pathogenicity, stress adaptation, and activating immune response of host. *Microbiol. Spectr.* 10(6), 2477–2422. doi: 10.1128/spectrum.02477-22

Memon, M. S., Kumar, P., and Iqbal, R. (2022). Meta deep learn leaf disease identification model for cotton crop. *Computers* 11 (7), 24. doi: 10.3390/computers11070102

Metzner, R., Eggert, A., van Dusschoten, D., Pflugfelder, D., Gerth, S., Schurr, U., et al. (2015). Direct comparison of MRI and X-ray CT technologies for 3D imaging of root systems in soil: potential and challenges for root trait quantification. *Plant Methods* 11, 11. doi: 10.1186/s13007-015-0060-z

Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01419

Poblete, T., Navas-Cortes, J. A., Camino, C., Calderon, R., Hornero, A., Gonzalez-Dugo, V., et al. (2021). Discriminating xylella fastidiosa from verticillium dahliae infections in olive trees using thermal- and hyperspectral-based plant traits. *Isprs J. Photogramm. Remote Sens.* 179, 133–144. doi: 10.1016/j.isprsjprs.2021.07.014

Ren, H., Li, X., Li, Y., Li, M., Sun, J., Wang, F., et al. (2022). Loss of function of VdDrs2, a P4-ATPase, impairs the toxin secretion and microsclerotia formation, and decreases the pathogenicity of verticillium dahliae. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.944364

Rogers, E. D., Monaenkova, D., Mijar, M., Nori, A., Goldman, D. I., and Benfey, P. N. (2016). X-Ray computed tomography reveals the response of root system architecture to soil texture. *Plant Physiol.* 171 (3), 2028–2040. doi: 10.1104/pp.16.00397

Santos-Rufo, A., and Rodriguez-Jurado, D. (2021). Unravelling the relationships among verticillium wilt, irrigation, and susceptible and tolerant olive cultivars. *Plant Pathol.* 70 (9), 2046–2061. doi: 10.1111/ppa.13442

Sayari, M., Dolatabadian, A., El-Shetehy, M., Rehal, P. K., and Daayf, F. (2022). Genome-based analysis of verticillium polyketide synthase gene clusters. *Biology-Basel* 11 (9), 30. doi: 10.3390/biology11091252

Schmittgen, S., Metzner, R., Van Dusschoten, D., Jansen, M., Fiorani, F., Jahnke, S., et al. (2015). Magnetic resonance imaging of sugar beet taproots in soil reveals growth reduction and morphological changes during foliar cercospora beticola infestation. *J. Exp. Bot.* 66 (18), 5543–5553. doi: 10.1093/jxb/erv109

Shaban, M., Miao, Y., Ullah, A., Khan, A. Q., Menghwar, H., Khan, A. H., et al. (2018). Physiological and molecular mechanism of defense in cotton against verticillium dahliae. *Plant Physiol. Biochem.* 125, 193–204. doi: 10.1016/j.plaphy.2018.02.011

Shen, C., Liu, L., Zhu, L., Kang, J., Wang, N., and Shao, L. (2020). High-throughput *in situ* root image segmentation based on the improved DeepLabv3+Method. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.576791

Sivakumar, P., Mohan, N. S. R., Kavya, P., and Teja, P. V. S. (2021). "Leaf disease identification: Enhanced cotton leaf disease identification using deep CNN models," in *IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT)*: IEEE. 22–26. doi: 10.1109/ICISSGT52025.2021.00016

Song, P., Wang, Z., Song, P., Yue, X., Bai, Y., and Feng, L. (2021). Evaluating the effect of aging process on the physicochemical characteristics of rice seeds by low field nuclear magnetic resonance and its imaging technique. *J. Cereal Sci.* 99, 8. doi: 10.1016/j.jcs.2021.103190

Song, P., Yue, X., Gu, Y., and Yang, T. (2022). Assessment of maize seed vigor under saline-alkali and drought stress based on low field nuclear magnetic resonance. *Biosyst. Eng.* 220, 135–145. doi: 10.1016/j.biosystemseng.2022.05.018

Wang, F., Rao, Y., Luo, Q., Jin, X., Jiang, Z., Zhang, W., et al (2022). Practical cucumber leaf disease recognition using improved Swin Transformer and small sample size. *Comput. Electron. Agr.* 199,. doi: 10.1016/j.compag.2022.107163

Wu, W., Ma, B.-L., and Whalen, J. K. (2018). "Enhancing rapeseed tolerance to heat and drought stresses in a changing climate: Perspectives for stress adaptation from root system architecture," in *Advances in agronomy*, vol. 151 . Ed. D. L. Sparks 151, 87–157. doi: 10.1016/bs.agron.2018.05.002

Yang, M., Huang, C., Kang, X., Qin, S., Ma, L., Wang, J., et al. (2022). Early monitoring of cotton verticillium wilt by leaf multiple "Symptom" characteristics. *Remote Sens.* 14 (20), 20. doi: 10.3390/rs14205241

Zhao, H., Wang, N., Sun, H., Zhu, L., Zhang, K., Zhang, Y., et al. (2022). RhizoPot platform: A high-throughput *in situ* root phenotyping platform with integrated hardware and software. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1004904

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 March 2023
DOI 10.3389/fpls.2023.1141153

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Milind B. Ratnaparkhe,
Indian Institute of Soybean Research (ICAR), India
Fan Shi,
Agriculture Victoria Research, Australia

*CORRESPONDENCE
Arti Singh
arti@iastate.edu
Baskar Ganapathysubramanian
baskarg@iastate.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 10 January 2023
ACCEPTED 28 February 2023
PUBLISHED 29 March 2023

CITATION
Young TJ, Jubery TZ, Carley CN, Carroll M, Sarkar S, Singh AK, Singh A and Ganapathysubramanian B (2023) "Canopy fingerprints" for characterizing three-dimensional point cloud data of soybean canopies. *Front. Plant Sci.* 14:1141153. doi: 10.3389/fpls.2023.1141153

COPYRIGHT
© 2023 Young, Jubery, Carley, Carroll, Sarkar, Singh, Singh and Ganapathysubramanian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

"Canopy fingerprints" for characterizing three-dimensional point cloud data of soybean canopies

Therin J. Young[1†], Talukder Z. Jubery[2†], Clayton N. Carley[3], Matthew Carroll[3], Soumik Sarkar[1,2], Asheesh K. Singh[3], Arti Singh[3*] and Baskar Ganapathysubramanian[1,2*]

[1]Department of Mechanical Engineering, Iowa State University, Ames, IA, United States, [2]Translational AI Center, Iowa State University, Ames, IA, United States, [3]Department of Agronomy, Iowa State University, Ames, IA, United States

Advances in imaging hardware allow high throughput capture of the detailed three-dimensional (3D) structure of plant canopies. The point cloud data is typically post-processed to extract coarse-scale geometric features (like volume, surface area, height, etc.) for downstream analysis. We extend feature extraction from 3D point cloud data to various additional features, which we denote as 'canopy fingerprints'. This is motivated by the successful application of the fingerprint concept for molecular fingerprints in chemistry applications and acoustic fingerprints in sound engineering applications. We developed an end-to-end pipeline to generate canopy fingerprints of a three-dimensional point cloud of soybean [*Glycine max* (L.) Merr.] canopies grown in hill plots captured by a terrestrial laser scanner (TLS). The pipeline includes noise removal, registration, and plot extraction, followed by the canopy fingerprint generation. The canopy fingerprints are generated by splitting the data into multiple sub-canopy scale components and extracting sub-canopy scale geometric features. The generated canopy fingerprints are interpretable and can assist in identifying patterns in a database of canopies, querying similar canopies, or identifying canopies with a certain shape. The framework can be extended to other modalities (for instance, hyperspectral point clouds) and tuned to find the most informative fingerprint representation for downstream tasks. These canopy fingerprints can aid in the utilization of canopy traits at previously unutilized scales, and therefore have applications in plant breeding and resilient crop production.

KEYWORDS

fingerprints, terrestrial laser scanner, 3D point cloud, phenotypic fingerprint, soybean

1 Introduction

Soybean [*Glycine max* (L.) Merr.] canopy characteristics indicate crop growth, development, and health among other characteristics. Canopy traits have traditionally focused on 2-dimensional (2D) features, which is useful in certain instances, for example, canopy coverage (Purcell, 2000), which has frequently been collected with drone high throughput phenotyping (Guo et al., 2021). With the advent of high-throughput crop and plant phenotyping (Araus and Cairns, 2014; Yang et al., 2020; Guo et al., 2021; Jubery et al., 2021; Singh A. K. et al, 2021; Singh D. P. et al., 2021), plant scientists have been able to conduct large scale and time-series investigations on canopy coverage. Additionally, researchers have shown automated or semi-automated extraction of canopy traits; for example, canopy features, including height, shape, color, and texture, can be used for plant stress and disease assessment, estimating total biomass, leaf chlorophyll, and leaf nitrogen (Shiraiwa and Sinclair, 1993; Hunt et al., 2005; Pydipati et al, 2006; Jubery et al., 2016; Bai et al., 2018; Parmley et al., 2019; Parmley et al., 2019). Canopy morphology features, such as shape and size, impact light interception ability, which directly factors into the potential yield equation (Metz et al, 1984; Koester et al., 2014). Canopy characteristics, including height, shape, size, and color, can vary among developmental stages, genotypes, and environments (Virdi et al., 2021). Quantifying the canopy plasticity of a genotype due to changing environmental conditions and variability or similarity among genotypes is valuable for plant breeding applications (Sadras and Slafer, 2012). However, a major hurdle towards effective and full utilization of canopy features is the relatively slow pace of advancement of three-dimensional (3D) canopy features, which provide a "real-world" set of information.

Historically, digital cameras, hyperspectral cameras, and LIDAR have been used to take images and create point clouds of plants (Walter et al., 2019; Herrero-Huerta et al., 2020; Chiozza et al., 2021a; Chang et al., 2022) which are then used to characterize plant traits, leading to a composite plant canopy. Often, methods such as structure from motion or tomographic reconstruction methods are needed to render the 3D point clouds for these traits (Vandenberghe et al, 2018; Storey, 2020). Widely utilized characterization approaches are based on hand-crafted geometric measures, such as plant height, length, breadth, height, area, and volume. Although these geometric features are simple to interpret, they often do not comprehensively represent the spatial variability, for instance, between sample heights and the intricacy of the canopies. Several studies used latent feature representation methods, such as Principal component analysis (PCA) and Neural Network (NN), to characterize the canopy (Gage et al., 2019; Ubbens et al., 2020). Although these features can be used to capture the inherent complexity of the canopies, they are challenging to comprehend since they are difficult to relate to real geometry with low interpretability. There is interest in developing more detailed yet interpretable phenotypic traits for characterizing the crop canopy. Interpretable features are crucial to develop field-testable hypotheses for plant scientists. Most interpretable approaches concentrate on composite characteristics and do not account for individual trait variations. An example approach that offers a middle ground between these two extremes is the elliptical Fourier transformation utilized to describe the complicated geometry of canopy structures (Jubery et al., 2016). However, the use of 3D point clouds can be more exhaustive and informative, motivating researchers to develop holistic phenotyping pipelines (end to end) as well as explore applications of the usage of information from these data. For example, 3D canopy generation has been successfully shown in wheat, *Triticum aestivum* (Paulus et al., 2013; Paulus et al., 2014), rice, *Oryza sativa* (Burgess et al., 2017; Zhu et al., 2018), and other crops (Vandenberghe et al, 2018). These are exciting developments; however, there is still information lacunae on the creation of informative multiscale traits from 3D point cloud data. In this context, non-agricultural disciplines have reported a concept of fingerprinting using point cloud data (Koutsoukas et al., 2014; Spannaus et al., 2021; Wang et al., 2021), but this is lacking in crop production and broader agriculture.

Fingerprinting is a technique for the multiscale characterization of an object by computing a set of unique local characteristics or patterns. Fingerprinting successfully generates unique representations for complex objects in chemistry, geometry, and acoustics (Cano et al., 2005; Capecchi et al, 2020). It was successfully used for the retrieval, recognition, and matching tasks within large molecular and 3D shape databases (Fontaine et al., 2007). Fingerprinting facilitates the representation of a complicated, memory-intensive 3D point cloud as a hierarchically computed, low-dimensional vector. This vector captures both the geometric and topological characteristics of 3D shapes. Computational approaches to computing fingerprinting for 3D objects are broadly based on spectral and non-spectral methods. Spectral approaches utilize the eigenvectors and eigenvalues, referred to as the spectrum, of the Laplace-Beltrami (LB) operator applied to 3D shapes (Reuter et al., 2005). The spectrum is independent of the object's representation, including the parameterization method and spatial position. Other techniques were developed from LB, for example, Shape-DNA and Global Point Signature (Wu et al, 2022). Probabilistic fingerprinting (Mitra et al., 2006) is a non-spectral fingerprinting technique. It is suitable for determining partial matching between 3D objects. Here, the objects were separated into overlapping patches, unique descriptors were generated for each patch, the descriptors were hashed, and a random subset of the hashed descriptors with a predetermined vector size was chosen as the probabilistic fingerprint. Similar min-hashing techniques (random subset selections) were used to get structural similarity in larger data based on chemistry (Probst, 2018). Hashing aids in the compression of the fingerprinting representation, but this cannot be decoded and is less interpretable. There are application examples of the fingerprinting concept; for example, a phenotypic fingerprint of a soybean canopy was proposed to represent the temporal variation of coarse-scale geometric features, including canopy height and plant length (Zhu et al., 2020b), and it was employed to capture temporal dynamics, identify genotypes with comparable growth signatures, etc. Further, canopy fingerprints enabled large-scale evaluation of the environmental constraints and disturbances that shape the 3D structure of forest canopies (Jucker, 2022). However, thus far, there is no work to define and develop crop canopy fingerprints capturing

multi-scale geometric features that could be evaluated and applied in the future in crop modeling, and genomic prediction (Jarquin et al, 2016; Shook et al, 2021a; Shook et al., 2021b), or breeding decisions. Fingerprints are distinctly unique from traditional canopy architecture as they encompass the entire global canopy, while architecture traits are a composite of limited individual traits assessed together.

The major contribution of this work is to develop an end-to-end non-spectral interpretable fingerprint generation pipeline for 3D point cloud data of field-grown row crops (Figure 1). The pipeline includes point cloud noise removal, registration, plot extraction, and fingerprint generation. We illustrate this approach using a large-scale field experiment through a diversity panel of soybeans. Specifically, we report the construction of canopy fingerprints in soybean using geometric and topological features of the 3D point cloud obtained by a Terrestrial laser scanner (TLS). This is accomplished with an end-to-end pipeline to generate canopy fingerprints of a three-dimensional point cloud of soybean, which is simple to use for feature extraction and utilization in a myriad of applications, including modeling, genomic prediction, ideotype breeding, and cultivar development. For example, the development of unique canopy fingerprints could enable faster and more efficient screening of genetic material for identifying canopy relationships with yield traits (Liu et al, 2016), biotic stress traits such as disease and insects (Pangga et al, 2011), how various canopy levels impact planting density, light interception, and photosynthesis (Feng et al., 2016), enable novel meta-GWAS (Shook et al., 2021a) or improve how crop modeling could predict the ideal canopy fingerprint (Rötter et al., 2015), or fingerprint ideotype, which could then be screened across core collections (Glaszmann et al., 2010) to narrow the pool of experimental genotypes *in silico* prior to *in vivo* evaluation.

2 Materials and methods

2.1 Laser scanner

The TLS used in this study was Trimble TX5 (Trimble Inc., Sunnyvale, CA, USA) (Figure 1A). It is a small and light device (240 mm x 200 mm x 100 mm in size and 5 kg in weight) that can perform measurements at speeds of 1 million points per second. The scanner collects data at a high angular resolution of 0.011 degrees, corresponding to a point spacing of 2 mm at a 10 m scanning range. The scanner emits a 3mm diameter and 905 nm wavelength laser beam and measures the distance between the scanner and the target using the phase-shift principle (Amann et al., 2001). The emitted laser beam is modulated at several frequencies, and the phase shift of all the returned modulations is assessed to increase the accuracy of the distance measurements while storing the intensity of the returned beam. The scanner covers a 360-degree x 300-degree field of view: 360 degrees on the vertical axis is achieved by rotating the scanner head, and a rotating mirror achieves 300 degrees on the horizontal axis. The scanner allows the acquisition of point clouds of 7.1 up to 710.7 million points (MP). The number of points corresponds to the resolution of the measurement. Additionally, the scanner has a built-in camera to capture RGB color values (up to 70 megapixels) and maps them to the corresponding point clouds.

2.2 Location, plant materials, and data collection

The experiment was done in a field frequently used for evaluating soybean iron deficiency chlorosis at Iowa State University's Agricultural Engineering/Agronomy Research Farm, IA, USA, at a latitude of 42.010 and a longitude of -93.735. Four hundred sixty-four soybean cultivars were included in this study. These accessions come from 35 countries and have crop maturities ranging from MG 0 to IV (Mourtzinis and Conley, 2017), along with variable seed weights and stem termination types. In May 2018, the cultivars were hand-planted in hill plots, three seeds per hill with 0.76 m spacing between each hill. Each plot consisted of a single hill replicated three times in the field, with each replication blocked together. No plants were thinned. Preparing a noise-free field is crucial for achieving accurate plant data. To minimize any interference from weeds, we conducted regular weeding at intervals throughout our study. Laser scanning was performed on all plots on the 9th and 10th of August. These scans were conducted 71 and 72 days after planting, and all plants had reached at least the

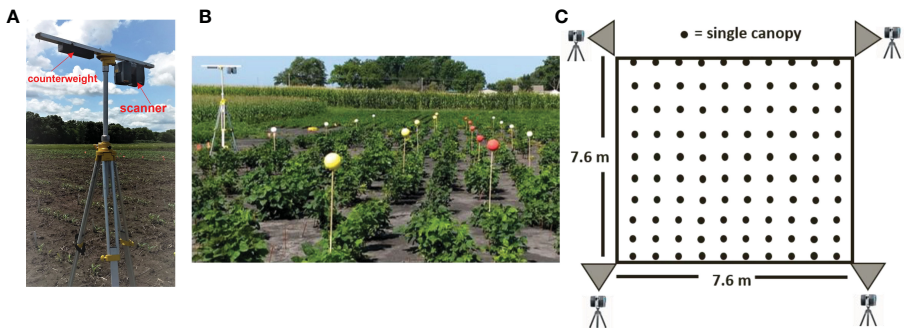


FIGURE 1
(A) Scanning platform: the scanner was mounted on a tripod in an inverted position with an extended bar and counterweight. **(B)** Placement of colored reference markers along the blocks. **(C)** Schematic of the scanning positions, block size, and canopy count per block.

reproductive R3 stage (Fehr et al., 1971) and were entirely opaque from the side, with no leaves visible from the opposite side.

The approximate size of the scanned field was 0.144 hectares (0.355 acres). During scanning, the field was divided into twenty-five 7.6 m x 7.6 m blocks, each containing 100 plots. The scanner was mounted on an 8 kg heavy-duty elevated tripod (Johnson Level, USA) at 2.1 m. For this height, the scanner can see the ground around the base of the farthest canopy within a block. This resulted in the typical scanner to ground distances between 2.1 to 11 m within the block. The device's scanning resolution was set at 0.5 (angular resolution 0.016 degree), and the expected point distance was 0.6 mm at 2.1 m and 3.1 mm at 11 m from the scanner.

To compensate for the scanner's field of view restriction of 150 degrees relative to the nadir, or lowest point under the observation lens, the scanner was mounted upside down using a 1.2 m-long bar, as illustrated in Figure 1A. This configuration allowed us to scan plots close to the tripod and position the scanner at the edge of each block. Scanning data was captured from four corners for each of the blocks: southwest (SW), southeast (SE), northeast (NE), and northwest (NW). The horizontal rotation limit of the laser scanner was set to 180 degrees, allowing two blocks to be scanned at once.

Before performing scans, Styrofoam spherical targets with a diameter of 0.127 m were placed within each block as reference markers to aid point cloud registration (Figure 1B), alignment, and plot identification. The spheres were painted yellow or red and mounted on 1.52 m-tall wooden dowels, which are 0.5 m taller than the expected maximum plant height. The dowels were manually pushed into the soil about 0.15 m deep. Each plot contained six reference markers. A white reference marker was placed at each corner of the block. The position of the reference markers was consistent across all scanned blocks.

Field experiments showed wind speeds to be lowest during the morning hours up until the early afternoon hours. When wind speeds exceeded 14.5 km h^{-1}, canopy movement exceeded the uncertainty acceptable for trait measurement. As a result, scanning took place between 9 a.m. and 2 p.m., or when wind speeds were 14.5 km h^{-1} or lower, to ensure the point cloud's quality was not compromised. While the optimal lighting condition for scanning is at noon, when sunlight is evenly distributed across the scanning area, field experiments demonstrated that overcast lighting also resulted in high-quality point cloud data. We avoided operating the scanner in the early mornings or late afternoons when direct or bright sunlight reflected from plant materials would cause laser signal saturation, resulting in erroneous points synonymous with glare in 2D photography.

Validation data consisting of plant height and canopy area were collected on August 8th. Plant height was recorded as the distance between the soil line at the base of the stem and the topmost leaf. The canopy area was defined as the visible area of the canopy from the nadir described in detail below.

On each plot, plant height was measured manually with a meter stick. The canopy area was measured on a subset of the plots as follows: First, a digital camera (Zenmuse X5 camera with a lens focal length of 45 mm)) mounted on a drone (Matrice 600 Pro) captured RGB images of the plots flown at 30m with 80% overlap and stitched together using Pix 4-D stitching software. We used an in-house Python script to extract individual plots from the stitched orthomosaic image, using the geolocation data obtained from the ground control points (GCP) and the RTK GPS mounted on the UAV. Next, we converted the images from RGB to the HSV color space, and the canopy was separated from the ground by applying a threshold to the Hue (H) color channel. We experimented with different threshold values for the Hue channel and found that the hue value worked best for our case. The canopy area was then calculated by determining the total number of non-zero (canopy) pixels and converting this value to m^2. To obtain the conversion factor from pixels to m^2, we measured a predefined marker in the images.

2.3 Point cloud processing pipeline

The pipeline was built using Python 3.7.3 and various other programs and packages, including Autodesk Recap 4.2.2.15, Cloud Compare 2.12.4 (Girardeau-Montaut n.d.), Open3D 0.11.2 (Zhou et al., 2018), and MATLAB 2019a. MATLAB and Cloud Compare were used *via* the command line interface and incorporated into the Python script *via* the Python subprocess library. The specific tasks carried out by these packages are depicted in Figure 2. The point cloud data was converted from FLS (Faro) to PCD (Point Cloud Library) format using Autodesk Recap Pro and Cloud Compare. Then, point cloud processing and trait extraction were performed

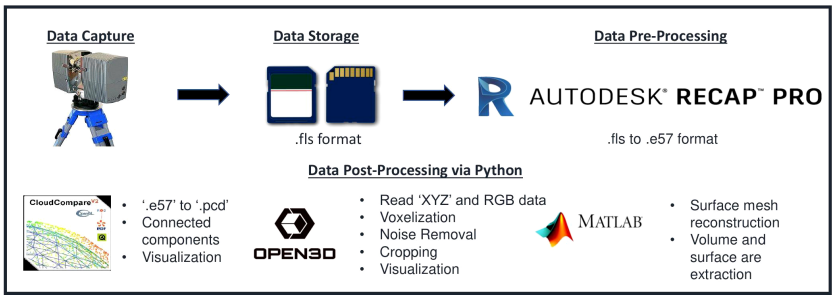


FIGURE 2
Data Processing Pipeline: Several applications were utilized in the pipeline. Autodesk recap pro was used to convert a scanner-vendor-specific file format to a generic one. CloudCompare and Open3D were employed for noise removal, voxelization, registration, and segmentation.

using Open3D, Cloud Compare, and MATLAB. This included cropping, voxelization, registration, noise removal, segmentation, and surface mesh reconstruction.

2.3.1 Point cloud file format conversion

The Trimble TX5 saves point cloud data in the FLS format, which is incompatible with the subsequent point cloud processing software. To ensure compatibility, the point cloud data format was converted from FLS to E57 using the Autodesk Recap Pro software. The E57 file format is a compact, vendor-independent format for storing point clouds, images, and metadata generated by 3D imaging systems such as laser scanners. Additionally, the E57 format retains the RGB component of the point cloud data. Finally, using Cloud Compare, the E57 files were converted to the PCD file format, which includes the Euclidean x, y, and z coordinates of each point and the RGB color value associated with each point.

2.3.2 Region of interest cropping

From the converted point cloud data, the region of interest, a block of the field, from each scan was automatically cropped out using the white-colored reference markers placed at the block's four corners. The markers were identified by separating points whose normalized R, G, and B color values are close to 1 and have a z coordinate (vertical direction, opposite of the gravity) value greater than 1 m. The z-value constraint was used to eliminate other white objects, such as orange and white plot stake identifiers. Then, the four white markers were identified as distinct objects using the connected components algorithm. Finally, the block was cropped out using the four markers' mean x and y coordinates (Figure 3A).

2.3.3 Homogenization

Due to the variable distances between the scanner and the plots, the point cloud density for a single plot captured from four different corners/perspectives varied (Figure 3B). This disparity can cause problems in mesh generation and skeletonization (Labussière et al., 2020; Xia et al., 2020). To reduce the disparity in point density, we voxelized the point clouds. The density of a point cloud is homogenized *via* uniform subsampling or voxel downsampling. We chose the latter method because it is more rigorous in ensuring uniform point distances and is invariant to the distribution of points within the sampling distance. It downsamples a point cloud uniformly using a regular voxel grid of 5 mm resolution. Briefly, voxels are used to group points, and each voxel generates an exact one point by averaging all points within an occupied voxel. Each point contains Euclidean x, y, and z coordinates and R, G, and B values.

2.3.4 Registration

Each block's four voxelized point clouds were co-registered and merged to form a single point cloud (Figure 3). The registration was carried out using the Cloudcompare 'Align' tool. We interactively identified a pair of color spheres in the point clouds, and then based on the center of the selected color spheres, the point clouds were aligned by rigid body transformation, ensuring the average root mean squared values of the distances between the paired points after registration is less than 0.01 m. When the preceding procedure failed to produce satisfactory results, we used the iterative closest point (ICP) algorithm to achieve fine registration. The tool can register up to two-point clouds in a single registration. As a result, three registrations were necessary to merge the four perspectives into a single cloud. The final registered point cloud contained duplicate points, and their density was inconsistent. Therefore, it was voxel-downsampled to restore the uniform point density in the registered point cloud.

2.3.5 Noise removal

Due to the so-called edge effect, in which a laser beam is partially intercepted at an object's edge, and the remaining beam travels further to collide with other objects or passes through the canopy, phase-shift lidar instruments, such as the Trimble TX5, are more prone to generate noisy spurious points *via* range averaging. Additionally, poor co-registration of point clouds and wind-driven movement of the plants can introduce noisy points.

A statistical-outlier-removal algorithm was used to remove noise in the registered point cloud (Figure 3C). The algorithm begins by calculating the average distance between each point and its (k) closest neighbors. Then it discards points whose average distance exceeds a predefined threshold, $\mu+\alpha\sigma$. μ and σ denote the mean and standard deviation of the average distances, respectively, and α is a parameter that can be tuned. The smaller the value of α, the more aggressive the point removal. By monitoring the deviation of a trait value (canopy height) for various combinations, the number of nearest neighbors, k and α, were selected.

2.3.6 Plot segmentation and ground .removal

The visible ground points between the plots were used to segment each plot and remove the ground. A plane ($z = f(x, y)$) was fitted to the registered point cloud, and the points above the fitted plane were retained (Figures 3). The plane passes through the middle of each plot, and the points above the plane comprise the top portion of the canopies. Each canopy top was labeled using the connected component algorithm, and each component's mean x and y coordinate was considered the plot's approximate center. The surrounding points within a square band of width 0.1 m and inner length 0.1 m are the faithful ground points for each plot. Finally, a plane was fitted through the ground points, and the points above the plane were considered the canopy.

2.4 Trait extraction

Height, volume, and surface area were extracted from the segmented canopy point clouds (Figure 3). The canopy height was determined by subtracting the minimum z-value in the canopy points from the mean z-value of the top 3% of canopy points. The choice of using the top 3% was based on a heuristic approach, as it yielded the closest agreement with the ground truth values (See Supplemental Material (S1)). To calculate volume and surface area, canopy points were bound into a tight 'watertight'

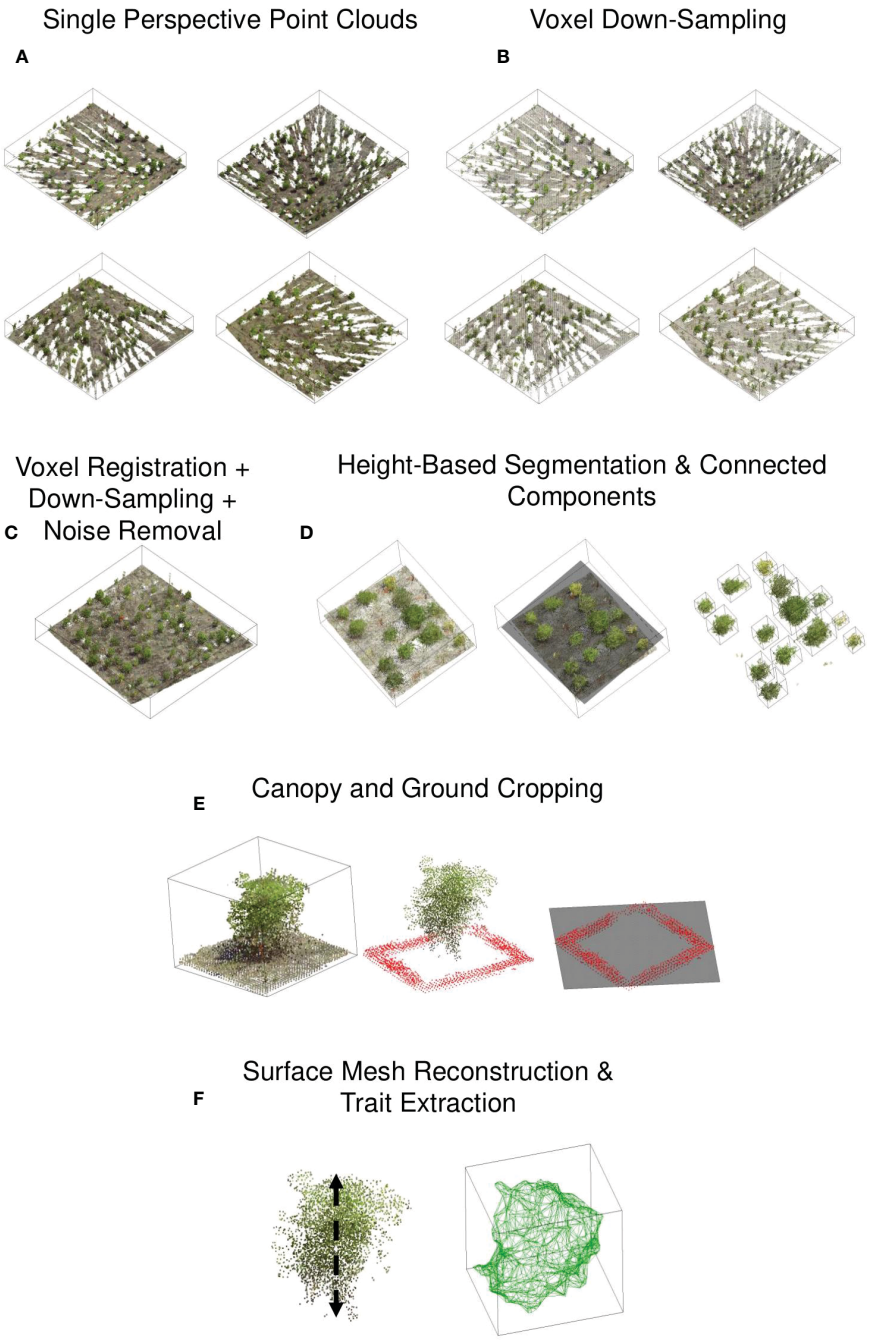


FIGURE 3
Point cloud processing pipeline: **(A)** The scanner captured the point cloud of a block at four corners of the block. The density of the point cloud is greater in proximity to the scanner. **(B)** Each point cloud was downsampled to reduce disparities in point cloud density. **(C)** The point clouds were registered, and the noise was removed. **(D)** Individual canopy detection was identified using height-based segmentation and connected components algorithm, **(E)** Ground point cloud was removed by identifying visible ground points around the canopy, **(F)** For the canopy point cloud, a triangular surface mesh was generated and the traits, including area, volume were computed.

triangular mesh using MATLAB's trisurf algorithm, and the volume and surface area were calculated using Python's trimesh library. Traits of the projected 2D outline of the point clouds were also extracted. First, the 3d point cloud of the canopy was projected onto the plane of interest. The projected points' boundary/contour was considered the 2D canopy outline. Area, aspect ratio, roundness, circularity, and solidity of the outline are defined as (Jubery et al., 2016):

- aspect ratio = major axis of the best-fit ellipse on the outline: minor axis of the best-fit ellipse on the outline; the ratio of the major to the minor axis of the best-fitted ellipse on the outline;
- roundness = 4 ∗ Area/(pi ∗ MajorAxis^2); it indicates the closeness of the shape of the outline to a circle;
- circularity = 4 ∗ pi ∗ Area/(Perimeter)$^{2;}$ it indicates the closeness of the form of the outline to a circle;

- solidity = Area/Convex Area; it is a measure of the compactness of the object.

2.5 Canopy fingerprints

Fingerprints are a way of representing complex physical objects mathematically. It can be used to illustrate various features on a local scale in a hierarchical and/or multi-scale way. Mathematical representation enables statistical or machine learning techniques to determine the similarity, signatures, and relationships between groups of objects. Here, we fingerprint a canopy by encoding it as a collection of sub-canopy-level features. For example, to fingerprint the shape of a canopy, we divide it into 2n+1 equally divided sections (sub-canopy). Here, we divide the canopy in the height direction into the 2n+1 sections. We then generate the signature of each sub-canopy using several geometric traits and normalize the traits concerning the traits of the center (n^{th}) sub-canopy. Finally, we represent normalized traits in a vector format to generate the fingerprint. Here 'n' is a tunable parameter that depends on the complexity of the canopies and intended downstream tasks involving the fingerprints (Figure 4).

3 Results and discussion

3.1 Parameters/conditions for TLS scanning and point cloud processing

Results showed that the point count (an indirect estimate of point density) of a canopy varies by its distance from the scanner, with as much as a 50% reduction when a canopy is close to the scanner versus when it is at the farthest possible distance from the scanner. However, this study circumvents the point density effect by registering multiple perspectives of the same scanned area. Thus, if the point density of a canopy near the scanner decreases as the scanner is moved farther away, the points lost can be recovered by scanning from a distinct perspective closer to the canopy (See Supplemental Material (S2)).

We used statistical outlier removal to reduce noise from the point cloud. Outlier selection is dependent on the values of two parameters: k, the number of neighbors, and alpha, the standard deviation ratio. We investigated the effect of 40 different combinations of these parameters on the extracted canopy height, including five values of the number of neighbors (k = 8, 16, 24, 32, 40) and eight levels of the standard deviation ratio.

The change in nearest neighbor parameter, k, from 8 to 16, average canopy height difference changed significantly after increasing (Figure 5). However, there were no significant changes in the average height difference for the remaining experiments (k = 32, k = 40). Additionally, following each experiment, visual analysis of the canopy point cloud revealed that most outliers were removed at k = 24. When k ≥ 24 and the standard deviation ratio was 0.075, no visually discernible changes in canopy structure occurred. Thus, the number of nearest neighbors, k, and the standard deviation ratio noise removal parameters were set at 24 and 0.075 for all noise removal tasks, respectively.

To evaluate the performance of co-registration, we determined the canopy top-view 2D projected area of a subsample of plants and compared it to the ground truth 2D projected area extracted from RGB images. By flattening the z values, the 3D point cloud of the canopy was projected onto the XY plane. A closed polyline represented the boundary/contour of the projected points, and its area was taken as the canopy area. An excellent agreement between the top-view canopy area and the ground truth area with R^2 = 0.826 (Figure 6).

3.2 Validation of the extracted canopy height

The results indicate that the extracted canopy height from the point cloud correlates with manual ground truth measurements taken on the same scanning day (Figure 7). Around 95% of the variation observed in extracted height values could be explained by a fitted linear least-squares regression model. Ground truth outliers in canopy height were defined as individuals with a Z-score greater than 2.5 compared to all samples' mean and standard deviation. Fewer than 2% of ground truth height measurements were considered outliers and were likely human errors in the collection. The deviation of heights from the manual measurements within the outliers ranged from 0.11 m to 0.43 m, with TLS measurements more frequently smaller than ground truth. After visualizing the outlier canopies' point cloud data, one explanation for the lower TLS height measurements is that occlusions between the measured canopy and neighboring canopies were not detected during data processing. However, manual height measurements of the canopies confirmed that the canopy segmentation and TLS height measurements were accurate (See Supplemental Material (S3)). Due to the low outlier rate (less

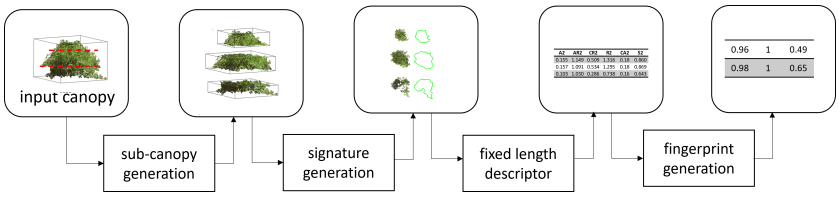


FIGURE 4
Overview of soybean canopy fingerprinting. Each canopy was subdivided into a predetermined number of sub-canopies, and the signature of each sub-canopy was extracted using several features, which were then arranged in a vector format and normalized with respect to the center sub-canopy features to represent the fingerprint of the canopy.

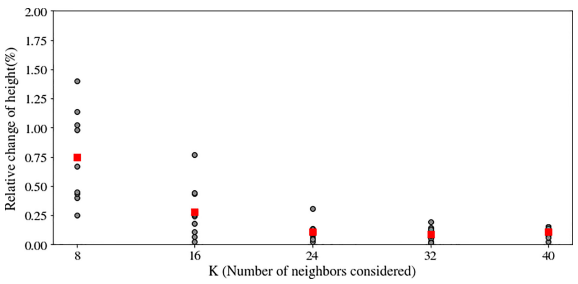


FIGURE 5
Effect of the number of neighbors (K) in the noise removal algorithm on the plant's height after noise removal. A large neighborhood size may eliminate both the actual canopy point cloud and noise, but a small neighborhood size may result in the preservation of noisy points. Optimal K was determined when increasing K had a negligible effect on plant height.

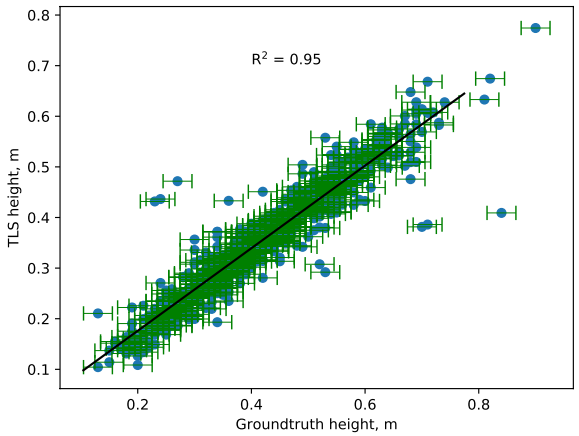


FIGURE 7
Comparison between manually measured canopy height (ground truth) and automatically measured canopy height from TLS-captured point cloud. The horizontal bar represents the measurement uncertainty associated with ground-truth data. The TLS-based plant height corresponds well with the actual plant height. Extreme outliers are believed to be the result of human mistakes.

than 2%) and high correlation (95%) between TLS and ground truth height data, TLS-based height extraction is a more robust method for canopy height measurement. This is useful because the TLS based method is automated.

3.3 Fingerprinting and implications

Typical drone and LiDAR 3D point clouds are often limited to a top-down view of the plant canopy due to collection limitations. Figure 8 depicts the canopy's conventional representation, as often shown from the whole plant perspective, and then shows a fingerprint perspective. The fingerprint representation was created by dividing the canopy into three and nine sub-canopies (2n+1, where n = 1 and 4). With these canopy fingerprints, we can find similar-looking canopies for a given canopy or a desired/given shape. Splitting these canopies into sub-canopies enables new opportunities for phenomics and further genomic assessment of cultivars. Traditionally, only the labor and time-extensive method of plant component partitioning would come close to this capability, but still lacked the ability for fingerprinting (Hintz and Albrecht, 1994; Raza et al., 2021). Sub-canopies paired with their fingerprints have the potential to further explore the unique relationships between certain fingerprint types or clusters with known canopy traits such as branching, leaf size, or leaf angle and their relationships with yield and yield component traits (Feng et al., 2018; Bianchi et al., 2020; Moro Rosso et al, 2021).

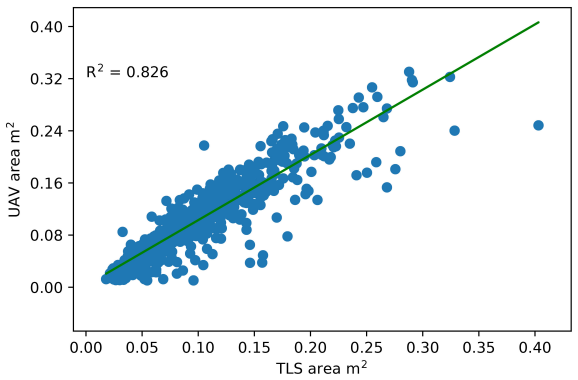


FIGURE 6
Comparison of the canopy area calculated from images captured by the UAV mount camera with the point clouds captured by the TLS. The TLS point cloud was projected onto a 2D XY plane (Top view), and the area of the closed contour around the projected point cloud was considered a canopy area.

With digital canopy fingerprints, we can now query a given canopy (Figure 9). The canopy point clouds database was converted to a searchable fingerprint database. To query a given canopy, a fingerprint of the canopy was generated and then compared with the existing database of fingerprints to identify the possible match. This capability could enable further in-depth development and exploration of the germplasm resources for ideotypes (Kokubun, 1988; Evangelista et al., 2021; Lukas et al., 2022; Roth et al., 2022). While this work evaluated a single time point, a more in-depth and temporal fingerprint could be developed to evaluate the canopy growth and development across time. These temporal fingerprints could open new insights into agronomic traits (Pfeiffer and Pilcher, 1987; Shiraiwa and Sinclair, 1993), diseases, and pesticide applications (Hanna et al., 2008; Sikora et al., 2014; Nagasubramanian et al., 2019; Viggers et al, 2022), or even develop fingerprint responses to abiotic stress and amendments (Frederick et al., 1991; Anda et al., 2021; Shivani Chiranjeevi et al, 2021; Bonds et al., 2022). An example of searching for potential ideotypes is shown in Figure 10, where some canopies with monotonically increasing mass at the top or the bottom are identified using the fingerprint representation. To find a canopy of the desired shape, we constructed a fingerprint of the desired shape and then determined which canopy fingerprints have the least Euclidean distance to the desired fingerprint. This further enables the exploration of canopy fingerprints *in silico* not only in relation to proposed ideotypes but also as a complement to crop modeling.

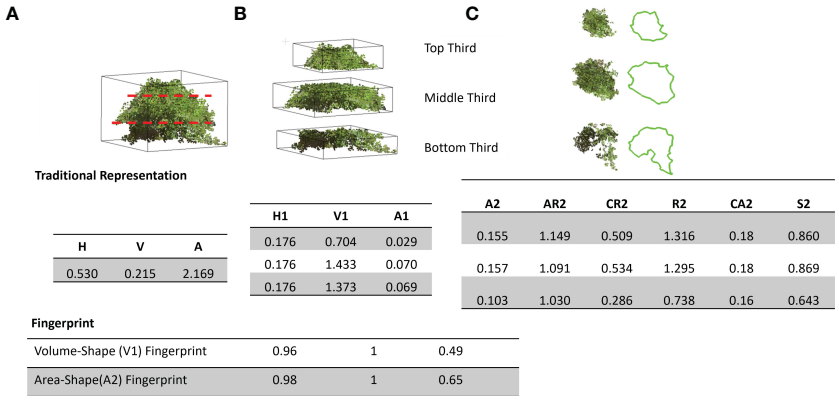


FIGURE 8
Traditional representation vs. canopy fingerprinting: Top panel: **(A)** Full canopy representation using height (H), volume (V), and surface area **(A, B)** Sub-canopy features including height (H1), volume(V1), Area(A1), **(C)** Sub-canopy 2D features including Area (A2), Aspect ratio (AR2), Circularity (CR2), Roundness (R2), Convex Area (CA2) and Solidity (S2), bottom panel: Fingerprint of the canopy shape Volume and Projected 2D Area (A2).

One of the core components of crop modeling is modeling the effect of light interception and radiation use efficiency of the canopy (Edwards et al, 2005; Singer et al., 2011; Zhu et al., 2020a). With canopy fingerprints integrated into a crop model, the theoretical evaluation of more genotypes in the models would be enabled, and stronger models could be developed and could also be expanded to explore environmental impacts and impacts on canopy fingerprints (Chiozza et al., 2021b; Krause et al., 2022).

An aspect of the fingerprinting to be further developed would be including the RGB data in the fingerprints. As apparent canopy color is related to soybean photosynthetic activity yield and plant health (Harrison et al, 1981; Rogovska et al, 2007; Naik et al., 2017; Yuan et al., 2019; Kaler et al., 2020; Rairdin et al., 2022). While the RGB data is already included within the voxels, additional work to evaluate the impact on fingerprint clustering due to color changes within each sublayer will be useful. Evaluating the color differences within each layer could provide an added trait assessment for radiation use efficiency relative to the amount of chlorophyll active in each canopy layer. Figure 11 shows PCA performed on the fingerprints and clustering. Each cluster's representative sample looks quite different and shows that the fingerprint can be used to pick diverse samples with the shape analysis alone. However, the inclusion of color data and additional layers, such as horizontal sub-layering, could further enable more detailed fingerprints for assessment and query while still reducing the computation load required for searching canopy fingerprint databases. These methods can parse out canopy features (through their fingerprinting) for a more informative representation of the canopy and the role of various organs throughout the canopy on desired traits, e.g., seed yield. This enables the discovery of new relationships between canopy and organ level features and their impact on yield and yield component traits.

While this work is focused on fingerprints assessed from TLS laser point clouds, the concept of canopy fingerprints could also be implemented with any technology capable of building a full canopy 3D point cloud, such as structured light, space-carving, or full canopy structure from motion (Nguyen et al., 2015; Zhou et al., 2019; Das Choudhury et al., 2020). While we focus on canopy fingerprints, further work should be done to evaluate whole plant fingerprints, especially root system architecture (RSA) fingerprints. While 2D imaging is routine for RSA traits (Falk et al, 2020b; Jubery et al.,

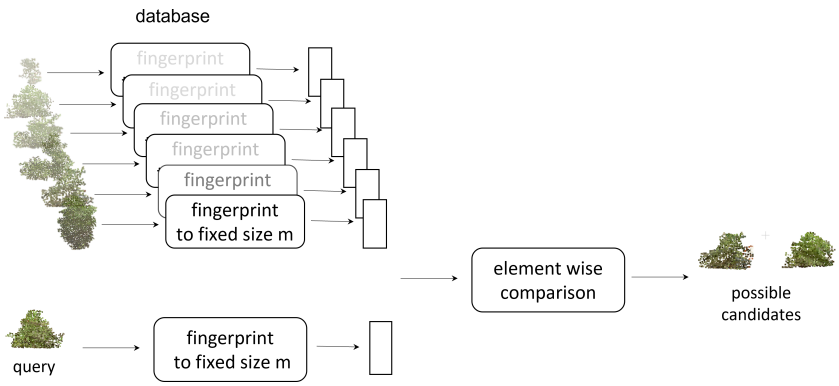


FIGURE 9
Query of a given canopy: the canopy point clouds database was converted to a fingerprint database. To query a given canopy, a fingerprint of the canopy was generated and then compared with the existing database of fingerprints to identify the possible match.

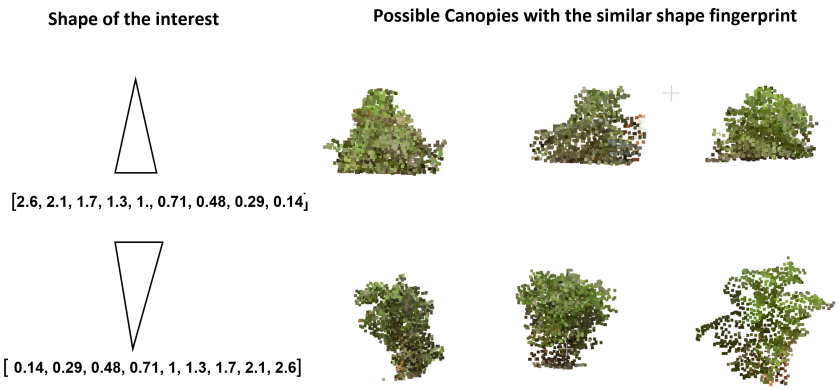


FIGURE 10
Canopies of given shapes (conical and inverted conical) were queried from the fingerprint database. The candidates look representative of the given shapes.

2021), there is tremendous interest in 3D imaging of root traits. RSA Fingerprints would further enable a whole plant analysis and efficient query system, and technology such as Xray-CT already enables dense 3D point clouds to be built of RSA (Gerth et al., 2021; Teramoto et al, 2021). Whole plant fingerprints could help meet the need for efficient RSA and canopy modeling, clustering, and assessment (Falk et al., 2020a; Carley et al., 2022a) while further exploring the root and shoot relationships to critical traits such as nodulation (Carley et al., 2022b). Irrespective of shoot or root fingerprints, there is tremendous potential for using this information to ID specific accessions and characterize germplasm collection (Azevedo Peixoto et al., 2017), cluster them based on their canopy features, develop relationships between agronomic, disease, or stress-induced traits, and modularize canopy features for their integration in trait development.

4 Conclusions

This study proposed an end-to-end fingerprint generation pipeline from a 3D point cloud of diverse soybean canopies grown on hill plots. The pipeline includes point cloud noise removal, registration, plot extraction, and fingerprint generation. Canopy fingerprinting is a generic and powerful approach to constructing interpretable, multi-scale, and/or hierarchical geometric traits from 3D point cloud data. This approach is a useful middle ground between conventional approaches of extracting coarse scale (i.e., full canopy scale) geometric features that may not comprehensively capture the spatial distribution of the canopy and the more recent approaches of directly compressing the point cloud data that produce difficult to interpret features. The generated fingerprints were used to query canopies of specific shapes to the group and identify similar canopies, which could be useful for future work in further identifying the relationships between canopy, agronomic traits, and yield relative to proposed ideotypes in varying climate scenarios. Canopy or whole plant fingerprinting could be used as a pre-classifier for a complete shape-based retrieval system. It could be used as a pseudo-leveler for self-supervised model training (REF) or useful in situations of limited annotation to train ML models (Kar et al., 2021; Nagasubramanian et al., 2021). Fingerprints could be added as a semantic tag (metadata) to the point cloud and can be queried instead of opening the data, and

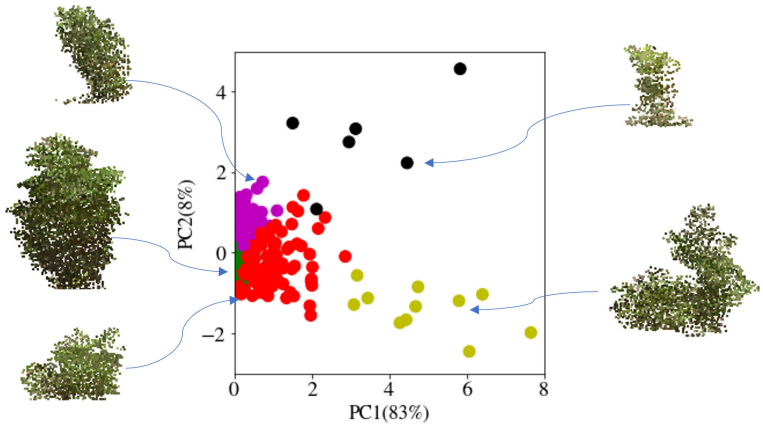


FIGURE 11
Identifying diversity in the canopy database: The fingerprints of the canopies were clustered and then visualized after dimensionality reduction using principal component analysis (PCA). The cluster-representative samples demonstrate the diversity of the canopies of each cluster.

can also be used for privacy-preserving deep models if data sharing is challenging (Cho et al., 2022). Fingerprinting also serves as a promising tool to store and quantify the inter-genotype or inter-environment variability. In combination with crop models and further development of voxel RGB data, these fingerprints could enable vast and rapid assessment of *in silico* genotypes for future experimentation in addition to the already improved searchability that fingerprint databases provide. As the fingerprint is based on simple sub-canopy level features, it has some limitations, and the proposed framework is sensitive to rigid transformations. If an upright plant becomes tilted, we get different fingerprint representations. However, this could be useful for estimating agronomic traits like lodging. Additionally, the vector of features as a function of plant height could be used for functional GWAS to explore putative loci with multi-scale canopy features. While our pipeline is built on TLS, future applications need to explore drone- and ground-based phenotyping (Gao et al., 2018; Guo et al., 2021; Riera et al., 2021). Plant phenotypic fingerprints serve as a novel opportunity to offer a diverse advantage to the future of high throughput phenotyping serving as a useful tool for data curation, cultivar selection, evaluation, and additional experimentation. Integration of canopy fingerprints with machine learning models can further advance the field of phenomics and cyber-agricultural systems (Singh et al., 2016; Singh et al., 2018; Singh A. K. et al., 2021).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

TY and TJ: Conceptualization; Data Collection; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing-original draft; Writing-review and editing. CC: Writing-original draft; Writing-review and editing. MC: Field experimentation; Data Collection; Writing-review and editing. SS: Funding acquisition; Resources; Supervision; Writing-review and editing. AKS: Funding acquisition; Resources; Experimentation; Writing-original draft; Writing-review and editing. AS: Conceptualization; Funding acquisition; Resources; Methodology; Supervision; Writing-review and editing. BG: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing-review and editing. All authors contributed to the article and approved the submitted version.

Funding

This project was supported by the Iowa Soybean Association, R.F. Baker Center for Plant Breeding, Plant Sciences Institute, USDA-NIFA Grants# 2017-67007-26151, 2017-67021-25965, AI Institute for Resilient Agriculture (USDA-NIFA #2021-67021-35329), COALESCE: COntext Aware LEarning for Sustainable CybEr-Agricultural Systems (NSF CPS Frontier # 1954556), FACT: A Scalable Cyber Ecosystem for Acquisition, Curation, and Analysis of Multispectral UAV Image Data (USDA-NIFA #2019-67021-29938), Smart Integrated Farm Network for Rural Agricultural Communities (SIRAC) (NSF S&CC #1952045), and USDA CRIS Project IOW04714. TY, CC, and MC were partially supported by the National Science Foundation under Grant No. DGE-1545453.

Acknowledgments

The authors thank all undergraduate and graduate students who helped with data collection, building registration markers, and field preparation for scanning. We also thank staff members, particularly Brian Scott and Will Doepke, for their assistance with experimentation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1141153/full#supplementary-material

SUPPLEMENTARY FIGURE 1
R-squared (R^2) values of the predicted TLS height against the ground truth values, with respect to varying % of top canopy points. The results showed that the R^2 value was highest when canopy height was calculated using the top 3% of the canopy points.

SUPPLEMENTARY FIGURE 2
Variation of point count of a canopy relative to its distance from the scanner.

SUPPLEMENTARY FIGURE 3
Examples of height measurement differences between TLS and Manual. The images depict interactive TLS (3D point cloud)-based height measurement within the CloudCompare software.

SUPPLEMENTARY FIGURE 4
This histogram displays the distribution of canopies based on the number of points they contain. The data pertains to 464 soybean cultivars including 450 plant introduction (PI) lines that were studied, representing a diverse range of maturities, seed weights, and stem terminations, and originating from 35 different countries. Data was obtained from one or two replicates per cultivar.

References

Amann, M.-C., Bosch, T. M., Lescure, M., Myllylae, R. A., and Rioux, M. (2001). Laser ranging: A critical review of unusual techniques for distance measurement. *Optical Eng.* 40 (January), 10–19. doi: 10.1117/1.1330700

Anda, A., Simon, B., Soós, G., Teixeira da Silva, J. A., and Menyhárt, L. (2021). Water stress modifies canopy light environment and qualitative and quantitative yield components in two soybean varieties. *Irrigation Sci.* 39 (5), 549–665. doi: 10.1007/s00271-021-00728-0

Araus, JoséL., and Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19 (1), 52–615. doi: 10.1016/j.tplants.2013.09.008

Azevedo Peixoto, L. de, Moellers, T. C., Zhang, J., Lorenz, A. J., Bhering, L. L., Beavis, W. D., et al. (2017). Leveraging genomic prediction to scan germplasm collection for crop improvement. *PloS One* 12 (6), e01791915. doi: 10.1371/journal.pone.0179191

Bai, G., Jenkins, S., Yuan, W., Graef, G. L., and Ge, Y. (2018). Field-based scoring of soybean iron deficiency chlorosis using RGB imaging and statistical learning. *Front. Plant Sci.* 9 (July), 1002. doi: 10.3389/fpls.2018.01002

Bianchi, J. S., Quijano, A., Gosparini, C. O., and Morandi, E. N. (2020). Changes in leaflet shape and seeds per pod modify crop growth parameters, canopy light environment, and yield components in soybean. *Crop J.* 8 (2), 351–645. doi: 10.1016/j.cj.2019.09.011

Bonds, D., Koziel, J. A., De, M., Chen, B., Singh, A. K., and Licht, M. A. (2022). Dataset documenting the interactions of biochar with manure, soil, and plants: Towards improved sustainability of animal and crop agriculture. *Brown Univ. Digest Addict. Theory Application: Data* 7 (3), 325. doi: 10.3390/data7030032

Burgess, A. J., Retkute, R., Herman, T., and Murchie, E. H. (2017). Exploring relationships between canopy architecture, light distribution, and photosynthesis in contrasting rice genotypes using 3D canopy reconstruction. *Front. Plant Sci.* 8 (May), 734. doi: 10.3389/fpls.2017.00734

Cano, P., Batlle, E., Kalker, T., and Haitsma, J. (2005). A review of audio fingerprinting. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* 41 (3), 271–845. doi: 10.1007/s11265-005-4151-3

Capecchi, A., Probst, D., and Reymond, J.-L. (2020). One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. *J. Cheminformatics* 12 (1), 435. doi: 10.1186/s13321-020-00445-4

Carley, C. N., Chen, G., Das, K. K., Delory, B. M., Dimitrova, A., Ding, Y., et al. (2022a). Root biology never sleeps: 11th symposium of the international society of root research (ISRR11) and the 9th international symposium on root development (Rooting2021), 24-28 may 2021. *New Phytol.* 235 (6), 2149–2154. doi: 10.1111/nph.18338

Carley, C. N., Zubrod, M., Dutta, S., and Singh, A. K. (2022b). Using machine learning enabled phenotyping to characterize nodulation in three early vegetative stages in soybean. *Crop Sci.* 63, 204–226. doi: 10.1101/2022.09.28.509969

Chang, S., Lee, U., Kim, J.-B., and Deuk Jo, Y. (2022). Application of 3D-volumetric analysis and hyperspectral imaging systems for investigation of heterosis and cytoplasmic effects in pepper. *Scientia Hortic.* 302 (August), 111150. doi: 10.1016/j.scienta.2022.111150

Chiozza, M., Parmley, K., Higgins, R. H., Singh, A. K., and Miguez, F. (2021). "Assessment of the G X e X m interaction effect on the soybean LAI-yield relationship using apsim model: Implications for high-throughput phenotyping," in *ASA, CSSA, SSSA International Annual Meeting* (ASA-CSSA-SSSA) Salt Lake City, UT. https://scisoc.confex.com/scisoc/2021am/meetingapp.cgi/Paper/138146.

Chiozza, M. V., Parmley, K. A., Higgins, R. H., Singh, A. K., and Miguez, F. E. (2021). Comparative prediction accuracy of hyperspectral bands for different soybean crop variables: From leaf area to seed composition. *Field Crops Res.* 271 (108260), 1082605. doi: 10.1016/j.fcr.2021.108260

Cho, M., Nagasubramanian, K., Singh, A. K., Singh, A., Ganapathysubramanian, B., Sarkar, S., et al. (2022) *Privacy-preserving deep models for plant stress phenotyping.* Available at: https://openreview.net/pdf?id=8JkuZSH487G.

Das Choudhury, S., Maturu, S., Samal, A., Stoerger, V., and Awada, T. (2020). Leveraging image analysis to compute 3D plant phenotypes based on voxel-grid plant reconstruction. *Front. Plant Sci.* 11 (December), 521431. doi: 10.3389/fpls.2020.521431

Edwards, J. T., Purcell, L. C., and Karcher, D. E. (2005). Soybean yield and biomass responses to increasing plant population among diverse maturity groups: II. light interception and utilization. *Crop Sci.* 45 (5), 1778–1785. doi: 10.2135/cropsci2004.0570

Evangelista, J. S. P. C., Peixoto, M. Antônio, Coelho, I. F., Alves, R. S., Silva, F. F. e., de Resende, M. D. V., et al. (2021). Environmental stratification and genotype recommendation toward the soybean ideotype: A Bayesian approach. *Crop Breed. Appl. Biotechnol.* 21 (1). doi: 10.1590/1984-70332021v21n1a11

Falk, K. G., Jubery, T. Z., Mirnezami, S. V., Parmley, K. A., Sarkar, S., Singh, A., et al. (2020a). Computer vision and machine learning enabled soybean root phenotyping pipeline. *Plant Methods* 161(55), 1–19. doi: 10.1186/s13007-019-0550-5

Falk, K. G., Jubery, T. Z., O'Rourke, J. A., Singh, A., Sarkar, S., Ganapathysubramanian, B., et al. (2020b). Soybean root system architecture trait study through genotypic, phenotypic, and shape-based clusters. *Plant Phenomics*.doi: 10.34133/2020/1925495

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, glycine max (L.) Merrill 1. *Crop Sci.* 11 (6), 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x

Feng, G., Luo, H., Zhang, Y., Gou, L., Yao, Y., Lin, Y., et al. (2016). Relationship between plant canopy characteristics and photosynthetic productivity in diverse cultivars of cotton (Gossypium hirsutum l.). *Crop J.* 4 (6), 499–508. doi: 10.1016/j.cj.2016.05.012

Feng, L., Raza, M. A., Li, Z., Chen, Y., Khalid, M. H. B., Du, J., et al. (2018). The influence of light intensity and leaf movement on photosynthesis characteristics and carbon balance of soybean. *Front. Plant Sci.* 9, 1952. doi: 10.3389/fpls.2018.01952

Fontaine, F., Bolton, E., Borodina, Y., and Bryant, S. H. (2007). Fast 3D shape screening of Large chemical databases through alignment-recycling. *Chem. Cent. J.* 1 (June), 12. doi: 10.1186/1752-153X-1-12

Frederick, J. R., Woolley, J. T., Hesketh, J. D., and Peters, D. B. (1991). Seed yield and agronomic traits of old and modern soybean cultivars under irrigation and soil water-deficit. *Field Crops Res.* 27 (1), 71–82. doi: 10.1016/0378-4290(91)90023-O

Gage, J. L., Richards, E., Lepak, N., Kaczmar, N., Soman, C., Chowdhary, G., et al. (2019). In-field whole-plant maize architecture characterized by subcanopy rovers and latent space phenotyping. *Plant Phenome J.* 2 (1), 1–115. doi: 10.2135/tppj2019.07.0011

Gao, T., Emadi, H., Saha, H., Zhang, J., Lofquist, A., Singh, A., et al. (2018). A novel multirobot system for plant phenotyping. *Robotics* 7 (4), 615. doi: 10.3390/robotics7040061

Gazzoni, DécioL., and Moscardi, Flávio *Effect of defoliation levels on recovery of leaf area, on yield and agronomic traits of soybeans1*. Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/44655/1/pab158-96.pdf (Accessed January 5, 2023).

Gerth, S., Claußen, J., Eggert, A., Wörlein, N., Waininger, M., Wittenberg, T., et al. (2021). Semiautomated 3D root segmentation and evaluation based on X-ray CT imagery. *Plant Phenomics (Washington D.C.)* 2021 (February), 8747930. doi: 10.34133/2021/8747930

Girardeau-Montaut, D. *CloudCompare - open source project*. Available at: https://www.danielgm.net/cc/ (Accessed December 11, 2022).

Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., and Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Biol.* 13 (2), 167–173. doi: 10.1016/j.pbi.2010.01.004

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., et al. (2021). UAS-based plant phenotyping for research and breeding applications. *Plant Phenomics (Washington D.C.)* 2021 (June), 9840192. doi: 10.34133/2021/9840192

Hanna, S. O., Conley, S. P., Shaner, G. E., and Santini, J. B. (2008). Fungicide application timing and row spacing effect on soybean canopy penetration and grain yield. *Agron. J.* 100 (5), 1488–1925. doi: 10.2134/agronj2007.0135

Harrison, S. A., Boerma, H. R., and Ashley, D. A. (1981). Heritability of canopy-apparent photosynthesis and its relationship to seed yield in soybeans 1. *Crop Sci.* 21 (2), 222–226. doi: 10.2135/cropsci1981.0011183X002100020004x

Herrero-Huerta, M., Eetu Puttonen, A. B., and Rainey, K. M. (2020). Canopy roughness: A new phenotypic trait to estimate aboveground biomass from unmanned aerial system. *Plant Phenomics*.doi: 10.34133/2020/6735967

Hintz, R. W., and Albrecht, K. A. (1994). Dry matter partitioning and forage nutritive value of soybean plant components. *Agron. J.* 86 (1), 59–625. doi: 10.2134/agronj1994.00021962008600010011x

Hunt, E.R., Cavigelli, M., Daughtry, C. S.T., Mcmurtrey, J. E., and Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precis. Agric.* 6 (4), 359–785. doi: 10.1007/s11119-005-2324-5

Jarquin, D., Specht, J., and Lorenz, A. (2016). Prospects of genomic prediction in the USDA soybean germplasm collection: Historical data creates robust models for enhancing selection of accessions. *G3* 6 (8), 2329–2415. doi: 10.1534/g3.116.031443

Jubery, T. Z., Carley, C. N., Singh, A., Sarkar, S., Ganapathysubramanian, B., and Singh, A. K. (2021). Using machine learning to develop a fully automated soybean nodule acquisition pipeline (SNAP). *Plant Phenomics (Washington D.C.)* 2021 (July), 9834746. doi: 10.34133/2021/9834746

Jubery, T. Z., Shook, J., Parmley, K., Zhang, J., Naik, H. S., Higgins, R., et al. (2016). Deploying Fourier coefficients to unravel soybean canopy diversity. *Front. Plant Sci.* 7, 2066. doi: 10.3389/fpls.2016.02066

Jucker, T. (2022). Deciphering the fingerprint of disturbance on the three-dimensional structure of the world's forests. *New Phytol.* 233 (2), 612–617. doi: 10.1111/nph.17729

Kaler, A. S., Abdel-Haleem, H., Fritschi, F. B., Gillman, J. D., Ray, J. D., Smith, J. R., et al. (2020). Genome-wide association mapping of dark green color index using a diverse panel of soybean accessions. *Sci. Rep.* 1–11. doi: 10.1038/s41598-020-62034-7

Kar, S., Nagasubramanian, K., Elango, D., Nair, A., Mueller, D. S., O'Neal, M. E., et al. (2021) *Self-supervised learning improves agricultural pest classification*. Available at: https://openreview.net/pdf?id=3ndIvP2491e.

Koester, R. P., Skoneczka, J. A., Cary, T. R., Diers, B. W., and Ainsworth, E. A. (2014). Historical gains in soybean (Glycine max merr.) seed yield are driven by linear

increases in light interception, energy conversion, and partitioning efficiencies. *J. Exp. Bot.* 65 (12), 3311–3215. doi: 10.1093/jxb/eru187

Kokubun, M. (1988). Design and examination of soybean ideotypes. *JARQ* 21 (4), 237–243.

Koutsoukas, A., Paricharak, S., Galloway, W. R.J., Spring, D. R., IJzerman, A. P., Glen, R. C., et al. (2014). How diverse are diversity assessment methods? a comparative analysis and benchmarking of molecular descriptor space. *J. Chem. Inf. Modeling.* 230–242. doi: 10.1021/ci400469u

Krause, M. D., Dias, K. O.G., Singh, A. K., and Beavis, W. D. (2022). Using Large soybean historical data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. *bioRxiv.* 2022–04. doi: 10.1101/2022.04.11.487885

Labussière, M., Laconte, J., and Pomerleau, François (2020). Geometry preserving sampling method based on spectral decomposition for Large-scale environments. *Front. Robotics AI* 7 (September), 572054. doi: 10.3389/frobt.2020.572054

Liu, T., Wang, Z., and Cai, T. (2016). Canopy apparent photosynthetic characteristics and yield of two spike-type wheat cultivars in response to row spacing under high plant density. *PloS One* 11 (2), e01485825. doi: 10.1371/journal.pone.0148582

Lukas, R., Barendregt, C., Bétrix, C.-A., Hund, A., and Walter, A. (2022). High-throughput field phenotyping of soybean: Spotting an ideotype. *Remote Sens. Environ.* 269, 112797. doi: 10.1016/j.rse.2021.112797

Metz, G. L., Green, D. E., and Shibles, R. M. (1984). Relationships between soybean yield in narrow rows and leaflet, canopy, and developmental characters 1. *Crop Sci.* 24 (3), 457–462. doi: 10.2135/cropsci1984.0011183X002400030006x

Mitra, N. J., Guibas, L. J., Giesen, J., and Pauly, M. (2006). "Probabilistic fingerprints for shapes," in *Symposium on geometry processing* (Germany: Eurographics Association, Goslar), 121–130.

Moro Rosso, L. H., de Borja Reis, AndréF., and Ciampitti, I. A. (2021). Vertical canopy profile and the impact of branches on soybean seed composition. *Front. Plant Sci.* 12 (September), 725767. doi: 10.3389/fpls.2021.725767

Mourtzinis, S., and Conley, S. P. (2017). Delineating soybean maturity groups across the united states. *Agron. J.* 109 (4), 1397–14035. doi: 10.2134/agronj2016.10.0581

Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., and Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* 15 (1), 1–10. doi: 10.1186/s13007-019-0479-8

Nagasubramanian, K., Jones, S., Singh, A., Singh, A., Sarkar, S., and Ganapathysubramanian, B. (2021). Plant phenotyping with limited annotation: Doing more with less. *Earth Space Sci. Open Archive.*doi: 10.1002/essoar.10508358.1

Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., et al. (2017). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods* 13 (April), 23. doi: 10.1186/s13007-017-0173-7

Nguyen, T. T., Slaughter, D. C., Max, N., Maloof, J. N., and Sinha, N. (2015). Structured light-based 3D reconstruction system for plants. *Sensors* 15 (8), 18587–16125. doi: 10.3390/s150818587

Pangga, I. B., Hanan, J., and Chakraborty, S. (2011). Pathogen dynamics in a crop canopy and their evolution under changing climate. *Plant Pathol.* 60 (1), 70–81. doi: 10.1111/j.1365-3059.2010.02408.x

Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., and Singh, A. K. (2019). Machine learning approach for prescriptive plant breeding. *Sci. Rep.* 9 (1), 171325. doi: 10.1038/s41598-019-53451-4

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., and Singh, A. K. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics (Washington D.C.)* 2019 (July), 5809404. doi: 10.34133/2019/5809404

Paulus, S., Behmann, J., Mahlein, A.-K., Plümer, L., and Kuhlmann, H. (2014). Low-cost 3D systems: Suitable tools for plant phenotyping. *Sensors* 14 (2), 3001–3185. doi: 10.3390/s140203001

Paulus, S., Dupuis, J., Mahlein, A.-K., and Kuhlmann, H. (2013). Surface feature based classification of plant organs from 3D laserscanned point clouds for plant phenotyping. *BMC Bioinf.* 14 (July), 238. doi: 10.1186/1471-2105-14-238

Pfeiffer, T. W., and Pilcher, D. (1987). Effect of early and late flowering on agronomic traits of soybean at different planting dates 1. *Crop Sci.* 27 (1), 108–112. doi: 10.2135/cropsci1987.0011183X002700010027x

Probst, D. (2018). And Jean-Louis reymond A probabilistic molecular fingerprint for big data settings. *J. Cheminformatics* 10 (1), 665. doi: 10.1186/s13321-018-0321-8

Purcell, L. C. (2000). Soybean canopy coverage and light interception measurements using digital imagery. *Crop Sci.* 40 (3), 834–837. doi: 10.2135/cropsci2000.403834x

Pydipati, R., Burks, T. F., and Lee, W. S. (2006). Identification of citrus disease using color texture features and discriminant analysis. *Comput. Electron. Agric.* 52 (1), 49–59. doi: 10.1016/j.compag.2006.01.004

Rairdin, A., Fotouhi, F., Zhang, J., Mueller, D. S., Ganapathysubramanian, B., Singh, A. K., et al. (2022). Deep learning-based phenotyping for genome wide association studies of sudden death syndrome in soybean. *Front. Plant Sci.* 13 (October), 966244. doi: 10.3389/fpls.2022.966244

Raza, M. A., Gul, H., Yang, F., Ahmed, M., and Yang, W. (2021). Growth rate, dry matter accumulation, and partitioning in soybean (Glycine max l.) in response to defoliation under high-rainfall conditions. *Plants* 10 (8), 1497. doi: 10.3390/plants10081497

Reuter, M., Wolter, F.-E., and Peinecke, N. (2005). "Laplace-Spectra as fingerprints for shape matching," in *Proceedings of the 2005 ACM symposium on solid and physical modeling* (New York, NY, USA: Association for Computing Machinery), 101–106.

Riera, L. G., Carroll, M. E., Zhang, Z., Shook, J. M., Ghosal, S., Gao, T., et al. (2021). Deep multiview image fusion for soybean yield estimation in breeding applications. *Plant Phenomics (Washington D.C.)* 2021 (June), 9846470. doi: 10.34133/2021/9846470

Rogovska, N. P., Blackmer, A. M., and Mallarino, A. P. (2007). Relationships between soybean yield, soil pH, and soil carbonate concentration. *Soil Sci. Soc. America J. Soil Sci. Soc. America* 71 (4), 1251–1565. doi: 10.2136/sssaj2006.0235

Roth, L., Barendregt, C., Bétrix, C.A., Hund, A., and Walter, A. (2022). High-throughput field phenotyping of soybean: Spotting an ideotype. *Remote Sensing Environ.* 269, 112797. doi: 10.1016/j.rse.2021.112797

Rötter, R. P., Tao, F., Höhn, J. G., and Palosuo, T. (2015). Use of crop simulation modelling to aid ideotype design of future cereal cultivars. *J. Exp. Bot.* 66 (12), 3463–3476. doi: 10.1093/jxb/erv098

Sadras, V. O., and Slafer, G. A. (2012). Environmental modulation of yield components in cereals: Heritabilities reveal a hierarchy of phenotypic plasticities. *Field Crops Res.* 127 (February), 215–224. doi: 10.1016/j.fcr.2011.11.014

Shiraiwa, T., and Sinclair, T. R. (1993). Distribution of nitrogen among leaves in soybean canopies. *Crop Sci.* 33 (4), 804–885. doi: 10.2135/cropsci1993.0011183X003300040035x

Shivani Chiranjeevi, T. Y., Jubery, T. Z., Nagasubramanian, K., Sarkar, S., Singh, A. K., Singh, A., et al. (2021). Exploring the use of 3D point cloud data for improved plant stress rating.

Shook, J. M., Lourenco, D., and Singh, A. K. (2021a). PATRIOT: A pipeline for tracing identity-by-Descent for chromosome segments to improve genomic prediction in self-pollinating crop species. *Front. Plant Sci.* 12 (676269), 2095. doi: 10.3389/fpls.2021.676269

Shook, J. M., Zhang, J., Jones, S. E., Singh, A., Diers, B. W., and Singh, A. K. (2021b). Meta-GWAS for quantitative trait loci identification in soybean. *G3 (Bethesda Md.)* 11 (7). doi: 10.1093/g3journal/jkab117

Sikora, E. J., Allen, T. W., Wise, K. A., Bergstrom, G., Bradley, C. A., Bond, J., et al. (2014). A coordinated effort to manage soybean rust in north America: A success story in soybean disease monitoring. *Plant Dis.* 98 (7), 864–875. doi: 10.1094/PDIS-02-14-0121-FE

Singer, J. W., Meek, D. W., Sauer, T. J., Prueger, J. H., and Hatfield, J. L. (2011). Variability of light interception and radiation use efficiency in maize and soybean. *Field Crops Res.* 121 (1), 147–525. doi: 10.1016/j.fcr.2010.12.007

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science* 21 (2), 110–124. doi: 10.1016/j.tplants.2015.10.015

Singh, A.K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends in Plant Science* 23(10), 883–898. doi: 10.1016/j.tplants.2018.07.004

Singh, A. K., Singh, A., Sarkar, S., Ganapathysubramanian, B., Schapaugh, W., Miguez, F. E., et al. (2021). "High-throughput phenotyping in soybean," in *High-throughput crop phenotyping.* Eds. J. Zhou and H. T. Nguyen (Cham: Springer International Publishing), 129–163.

Singh, D. P., Singh, A. K., and Singh, A. (2021). *Plant breeding and cultivar development* (San Diego, CA: Academic Press).

Spannaus, A., Law, K. J.H., Luszczek, P., Nasrin, F., Micucci, C. P., Liaw, P. K., et al. (2021). Materials fingerprinting classification. *Comput. Phys. Commun.* 266 (September), 108019. doi: 10.1016/j.cpc.2021.108019

Storey, E. A. (2020). Mapping plant growth forms using structure-from-Motion data combined with spectral image derivatives. *Remote Sens. Lett.* 11 (5), 426–435. doi: 10.1080/2150704X.2020.1730467

Teramoto, S., Tanabata, T., and Uga, Y. (2021). RSAtrace3D: Robust vectorization software for measuring monocot root system architecture. *BMC Plant Biol.* 21 (1), 3985. doi: 10.1186/s12870-021-03161-9

Ubbens, J., Cieslak, M., Prusinkiewicz, P., Parkin, I., Ebersbach, J., and Stavness, I. (2020). Latent space phenotyping: Automatic image-based phenotyping for treatment studies. *Plant Phenomics (Washington D.C.)* 2020 (January), 5801869. doi: 10.34133/2020/5801869

Vandenberghe, B., Depuydt, S., and Van Messem, A. (2018). How to make sense of 3D representations for plant phenotyping: A compendium of processing and analysis techniques. doi: 10.31219/osf.io/r84mk

Viggers, J. N., Kandel, Y. R., and Mueller, D. S. (2022). The impact of fungicide application method on soybean canopy coverage, frogeye leaf spot, and yield. *Plant Health Progress.*doi: 10.1094/php-03-22-0024-rs

Virdi, K. S., Sreeka, S., Dobbels, A., Haaning, A., Jarquin, D., Stupar, R. M., et al. (2021). Branch angle and leaflet shape are associated with canopy coverage in soybean. *Res. Square.*doi: 10.21203/rs.3.rs-806530/v1

Walter, J. D.C., Edwards, J., McDonald, G., and Kuchel, H. (2019). Estimating biomass and canopy height with LiDAR for field crop breeding. *Front. Plant Sci.* 10 (September), 1145. doi: 10.3389/fpls.2019.01145

Wang, L., Zhao, Lu, Liu, X., Fu, J., and Zhang, A. (2021). SepPCNET: Deeping learning on a 3D surface electrostatic potential point cloud for enhanced toxicity classification and its application to suspected environmental estrogens. *Environ. Sci. Technol.* 9958–9967. doi: 10.1021/acs.est.1c01228

Wu, Y., Wu, T., and Yau, S.-T. (2022). Surface eigenvalues with lattice-based approximation in comparison with analytical solution. doi: 10.48550/arXiv.2203.03603

Xia, S., Chen, D., Wang, R., Li, J., and Zhang, X. (2020). Geometric primitives in LiDAR point clouds: A review. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 13, 685–707. doi: 10.1109/JSTARS.2020.2969119

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* 13 (2), 187–2145. doi: 10.1016/j.molp.2020.01.008

Yuan, W., Wijewardane, N. K., Jenkins, S., Bai, G., Ge, Y., and Graef, G. L. (2019). Early prediction of soybean traits through color and texture features of canopy RGB imagery. *Sci. Rep.* 9 (1), 140895. doi: 10.1038/s41598-019-50480-x

Zhou, J., Fu, X., Zhou, S., Zhou, J., Ye, H., and Nguyen, H. T. (2019). Automated segmentation of soybean plants from 3D point cloud using machine learning. *Comput. Electron. Agric.* 162 (July), 143–153. doi: 10.1016/j.compag.2019.04.014

Zhou, Q.-Y., Park, J., and Koltun, V. (2018). Open3D: A modern library for 3D data processing. doi: 10.48550/arXiv.1801.09847

Zhu, B., Liu, F., Xie, Z., Guo, Y., Li, B., and Ma, Y. (2020a). Quantification of light interception within image-based 3-d reconstruction of sole and intercropped canopies over the entire growth season. *Ann. Bot.* 126 (4), 701–125. doi: 10.1093/aob/mcaa046

Zhu, R., Sun, K., Yan, Z., Yan, X., Yu, J., Shi, J., et al. (2020b). Analysing the phenotype development of soybean plants using low-cost 3D reconstruction. *Sci. Rep.* 10 (1). doi: 10.1038/s41598-020-63720-2

Zhu, F., Thapa, S., Gao, T., Ge, Y., Walia, H., and Yu, H. (2018). "3D reconstruction of plant leaves for high-throughput phenotyping," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA. 2018, 4285–4293. doi: 10.1109/BigData.2018.8622428

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 April 2023
DOI 10.3389/fpls.2023.1155722

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service, United States Department of Agriculture (USDA), United States

REVIEWED BY
Kusumiyati Kusumiyati,
Padjadjaran University, Indonesia
Guy Samson,
Université du Québec à Trois-Rivières, Canada

*CORRESPONDENCE
Changhyeon Kim
kim.9067@osu.edu

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 31 January 2023
ACCEPTED 21 March 2023
PUBLISHED 03 April 2023

CITATION
Kim C and van Iersel MW (2023) Image-based phenotyping to estimate anthocyanin concentrations in lettuce. *Front. Plant Sci.* 14:1155722. doi: 10.3389/fpls.2023.1155722

COPYRIGHT
© 2023 Kim and van Iersel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Image-based phenotyping to estimate anthocyanin concentrations in lettuce

Changhyeon Kim[1]* and Marc W. van Iersel[2]

[1]Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH, United States, [2]Department of Horticulture, The University of Georgia, Athens, GA, United States

Anthocyanins provide blue, red, and purple color to fruits, vegetables, and flowers. Due to their benefits for human health and aesthetic appeal, anthocyanin content in crops affects consumer preference. Rapid, low-cost, and non-destructive phenotyping of anthocyanins is not well developed. Here, we introduce the normalized difference anthocyanin index (NDAI), which is based on the optical properties of anthocyanins: high absorptance in the green and low absorptance in the red part of the spectrum. NDAI is determined as $(I_{red} - I_{green})/(I_{red} + I_{green})$, where I is the pixel intensity, a measure of reflectance. To test NDAI, leaf discs of two red lettuce (*Lactuca sativa*) cultivars 'Rouxai' and 'Teodore' with wide range of anthocyanin concentrations were imaged using a multispectral imaging system and the red and green images were used to calculate NDAI. NDAI and other commonly used indices for anthocyanin quantification were evaluated by comparing to with the measured anthocyanin concentration (n = 50). Statistical results showed that NDAI has advantages over other indices in terms of prediction of anthocyanin concentrations. Canopy NDAI, obtained using multispectral canopy imaging, was correlated (n = 108, R^2 = 0.73) with the anthocyanin concentrations of the top canopy layer, which is visible in the images. Comparison of canopy NDAI from multispectral images and RGB images acquired using a Linux-based microcomputer with color camera, showed similar results in the prediction of anthocyanin concentration. Thus, a low-cost microcomputer with a camera can be used to build an automated phenotyping system for anthocyanin content.

KEYWORDS

anthocyanins, remote sensing, anthocyanin index, non-destructive measurement, low-cost plant phenotyping, controlled environment agriculture (CEA)

1 Introduction

Anthocyanins are water-soluble pigments that provide red, purple, or blue color to leaves, fruits, and flowers. Anthocyanins accumulate in response to various abiotic and biotic stresses and provide protection to plants against these stressors (Boldt et al., 2014; Liu et al., 2018). Benefits of anthocyanins for human health include anti-cancer activity and alleviating cardiovascular disease and diabetes (Khoo et al., 2017). Due to their benefits and

aesthetic appeal, anthocyanins can influence consumer preference. Therefore, rapid phenotyping for anthocyanin content is important in horticultural production and breeding programs, as well as in ecophysiological studies.

Non-destructive estimation of traits using imaging is an ideal approach to phenotyping. Image-based plant phenotyping quantifies reflected light from plants in a non-destructive manner, while the reflected light has unique spectral responses to plant pigments. Multispectral or hyperspectral imaging sensors detect reflected light from plants, and that information is stored as pixels with spatial and quantitative information regarding reflected light intensity and color. Image-based phenotyping can provide reliable data to characterize traits of interest in a non-destructive, rapid, and high-throughput manner, without sampling bias (Del Valle et al., 2018; Costa et al., 2019). These advantages led to an emergence of image-based phenotyping in the early 2010s and many researchers have adopted such systems to screen for traits of interest (Das Choudhury et al., 2019; Maes and Steppe, 2019).

Reflectance imaging can be applied at various scales, including satellites, and plant phenotyping has benefitted from prior work using satellite imaging. Commonly, reflectance indices are calculated using normalization equations, which constrain the index values between -1 and 1. Common indices to quantify vegetation cover, plant health, or physiological status include the normalized difference vegetation index (NDVI) and photochemical reflectance index (PRI) (Rouse et al., 1974; Gamon et al., 1997). Such indices are derived from changes in reflectance based on canopy size or the nutritional and/or physiological status of plants.

Prior work resulted in indices to estimate anthocyanin concentrations: the red to green ratio (Gamon and Surfus, 1999), anthocyanin reflectance index (Gitelson et al., 2001), and modified anthocyanin reflectance index (Gitelson et al., 2006). Bayle et al. (2019) reported the normalized anthocyanin reflectance index (NARI), which is a modified version of the anthocyanin reflectance index. These indices are based on the higher absorptance (or lower reflectance) in the green part of the spectrum (500 – 550 nm) of plants with higher anthocyanin concentrations (Neill and Gould, 2000; Gitelson et al., 2001; Merzlyak et al., 2008). These indices also include red or red-edge reflectance to adjust for the presence of chlorophylls. Most of these indices use reflectance in the red edge (700 - 705 nm) for chlorophyll corrections, due to the strong correlation between the reflectance in the red edge and chlorophyll concentration (Gitelson et al., 1996). At the same time, Gitelson et al. (2001) and Gamon and Surfus (1999) also reported a correlation between chlorophyll concentrations and reflectance in a wide range of the red spectrum (600 - 700 nm). Therefore, anthocyanin predictions may be achieved using reflectance in a wide range of the green and red spectrum, which can be easily acquired by low-cost color (RGB) imaging. However, little is known about the feasibility of predicting anthocyanin concentration using reflectance indices with spectral images acquired by low-cost imaging systems.

With the advance in computing power and better imaging sensors, image-based phenotyping tends to use hyperspectral imaging and machine learning algorithms to process high-throughput data. At the same time, the development of simple indices has become less popular. It is understandable that machine learning approaches have become popular in plant science, because the amount of data generated by hyperspectral imaging system is large, so identification of traits of interest based on relatively simple analyses may not be feasible. However, machine learning may not always provide ideal solutions and may result in errors when the conditions during image acquisition differ from those under which machine learning algorithms were developed. Additionally, machine learning-based models cannot provide underlying physiological meaning to these models.

Using hyperspectral imaging systems to quantify phenotypic traits of interest can be useful, but is not always necessary. Hyperspectral imaging can be expensive and simple alternative approaches may make imaging accessible to many more scientists, as well as horticultural producers. For example, image-based phenotyping of anthocyanin concentrations only requires information in the red and green wavebands based on the optical characteristics of anthocyanins and chlorophyll, so hyperspectral imaging may not be necessary. Furthermore, color imaging provides quantitative information on reflectance in the red, green, and blue wavebands, which may satisfy the requirements for anthocyanin prediction. Therefore, we hypothesized that simple color cameras or multispectral imaging systems can provide quantitative information regarding anthocyanin concentrations.

Here, we introduce a new reflectance index, the normalized difference anthocyanin index (NDAI), for predicting anthocyanin concentration using the images of anthocyanin-rich lettuce cultivars. We evaluated the performance of NDAI by comparing it with other anthocyanin indices in use. We developed a Python script to calculate these indices using multispectral images of leaf discs from two red lettuce cultivars, 'Rouxai' and 'Teodore' and compared those values to measured anthocyanin concentrations. We also tested whether a low-cost RGB camera, connected to a Raspberry Pi microcomputer can be used to determine the NDAI. Our objectives were (1) to evaluate the performance of different indices, including NDAI, in terms of the prediction of the leaf disc anthocyanin concentration, (2) to validate the best performing index for prediction of anthocyanin concentration at the whole plant scale, and (3) to develop a low-cost RGB imaging system that can predict anthocyanin concentrations.

2 Materials and methods

2.1 Plant materials

Red-leaf lettuce (*Lactuca sativa*) cultivars 'Rouxai' and 'Teodore' were grown in a walk-in growth chamber from April 14 to May 14, 2021. Seeds were planted in 10 cm square pots containing a soilless substrate (Fafard 2P Mix; Sun Gro Horticulture, Agawam, MA, USA). The plants were subirrigated biweekly with a water-soluble fertilizer (15N-2.2P-12.5K; Peters Excel 15-15-15 Cal-Mag special Fertilizer, ICL Fertilizers, Dublin, OH, USA). The approximate environmental conditions of the walk-in chamber were a photoperiod of 16 hours, photosynthetic photon flux density (*PPFD*) of 250 μmol m^{-2} s^{-1}, daily light integral of 14.4

mol m^{-2} d^{-1}, a temperature of 20 °C, a vapor pressure deficit of 1.2 kPa, and CO_2 concentration of 800 µmol mol^{-1}. After 30 days, the plants were transferred to growth chambers with temperatures of 4, 12, and 20 °C for 0, 12, 24, or 36 hours to induce a wide range of anthocyanin concentrations. The other environmental conditions in these growth chambers were similar to those of the walk-in cooler. Each combination of the temperature, exposure time, and cultivar was repeated three times with three biological replications.

2.2 Sample collection and multispectral imaging

A randomly-selected plant from each exposure time and temperature combination was used for leaf disc imaging and anthocyanin extraction. Two positions on leaves near the top of the plant were selected based on visual assessment of homogeneous anthocyanin distribution. A cardboard piece with a hole with 1-cm radius was clamped on the leaves. The 1-cm exposed leaf discs were imaged using a commercial imaging system (TopView, ARIS, Eindhoven, The Netherlands). To acquire spectral images, the monochrome camera of the imaging system captures one monochrome image at a time under illumination of a monochrome light-emitting diodes (LED) and repeats the process with sequential illumination of seven wavelengths of LED (450, 516, 593, 625, 664, 730, and 861 nm peaks) (Figure 1A). We only used the images taken under the green (peak at 516 nm), red (peak at 664 nm), and near infrared (NIR, peak at 861 nm) LEDs (Figure 1B), following previous studies of anthocyanin prediction (Gamon and Surfus, 1999; Gitelson et al., 2006; Gitelson et al., 2009; Steele et al., 2009). The full width at half maximum (FWHM) of the green, red, and NIR LEDs was 40, 25 and 28 nm, respectively. During the NIR imaging, the imaging system could not achieve perfect focus. Therefore, image indices using NIR image resulted in a relatively poor spatial resolution (Spectrum of NIR in Figure 1B).

Before taking plant images, a middle-gray card (50% reflection at all wavebands) was imaged as a reference to calibrate exposure times of each LED. We measured the pixel intensity of the gray card using Image J software (NIH, Bethesda, MD, USA). Since the middle-gray card should result in an intensity of 127 in a scale of 255 (8-bit resolution of the images), we adjusted the exposure time based on the ratio between 127 and the measured pixel intensity. The calibration with the gray card ensures that images taken under different colors of LEDs are directly comparable and accurate measures of reflectance.

After the multispectral imaging, the 1-cm radius leaf discs were dissected and frozen in liquid nitrogen before anthocyanin extraction and quantification, following a modified protocol of Lee et al. (2005). A leaf disc was weighed and ground with liquid nitrogen. The ground sample was then precipitated in 5 ml of methanol with 2% HCl. After centrifugation, 0.5 ml of the supernatant was transferred to a buffer solution containing 2 ml of 0.025 M KCl (pH 1.0) and 2 ml of 0.4 M sodium acetate (pH 4.5). Absorbance of the solution at 520 and 720 nm was measured using a spectrophotometer (GENESYS 10S UV-Vis spectrophotometer, Thermo Scientific™, Waltham, MA, USA) with three subsamples per the solution. The measured absorbance and the fresh weight (FW) were then used to compute the anthocyanin concentration of the leaf disc per unit FW. A total of 50 leaf discs were used for anthocyanin extraction to evaluate the performance of NDAI and other indices from literatures regarding prediction of anthocyanin concentration.

To test feasibility of predicting anthocyanin concentration in canopy layer using a best performing index in the leaf discs, the intact canopy of 108 lettuce plants was also imaged using the TopView system and used for anthocyanin extraction. Two randomly selected plants from each exposure time and temperature combination were imaged. Since these images only capture the top part of the plant canopy, we collected only the leaves from the top layer of the canopy for anthocyanin extraction

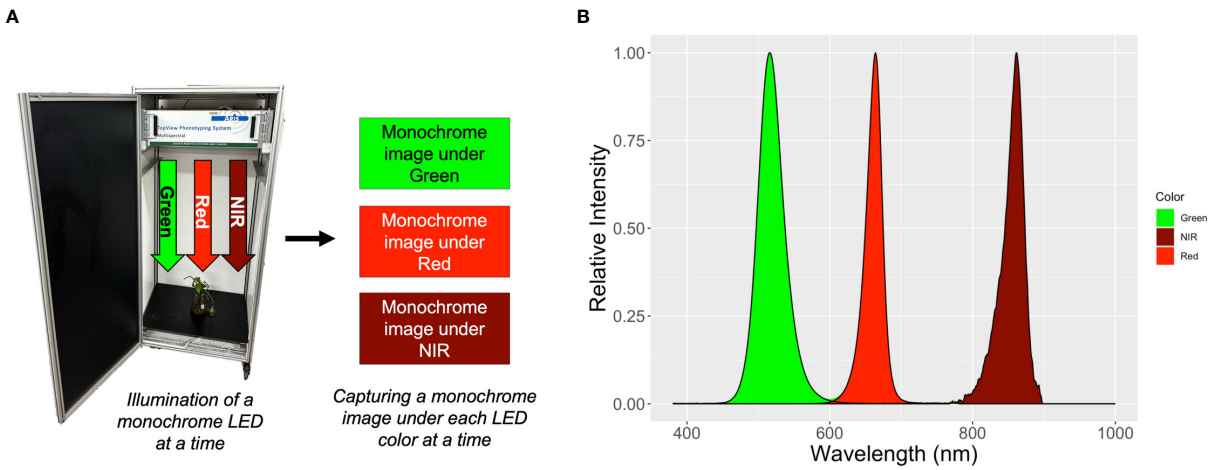


FIGURE 1
Schematic representation of the acquisition process of green, red, and near-infrared (NIR) spectral images by the TopView multispectral imaging system **(A)** and the normalized spectrum of the LEDs of the system used for capturing each monochrome image **(B)**. The peak and full width at half maximum of green, red, and NIR LEDs were 516 and 40 nm, 664 and 25 nm, and 861 and 28 nm, respectively.

following the same protocol. The leaves were ground and homogenized, and approximately 500 mg of the homogenized leaf tissue was used for the extraction.

2.3 Anthocyanin index comparisons

The new Normalized Difference Anthocyanin Index (NDAI) is defined as:

$$NDAI = (I_{red} - I_{green})/(I_{red} + I_{green}) \quad (1)$$

where I is the pixel intensity, a measure of reflectance, and the subscript indicates the color of the image, while peak and FWHM of the light color vary depending on the light spectrum during imaging. This index uses the same basic equation as the normalized difference vegetation index (Rouse et al., 1974) and photochemical reflectance index (Gamon et al., 1997), which are commonly used normalized indices in remote sensing. We used the green and red wavebands based on the optical properties of anthocyanins *in situ*; higher anthocyanin concentrations result in higher absorptance in the green part of the spectrum, while the absorptance in the red part of the spectrum is related to chlorophyll content (Lichtenthaler, 1987; Gitelson et al., 2001). Therefore, we speculate that NDAI has an advantage over indices that do not use the red part of the spectrum by accounting for chlorophyll content. Indices that use NIR, instead of red, cannot do so.

Bayle et al. (2019) developed the normalized anthocyanin reflectance index (NARI):

$$NARI = ((R_{550})^{-1} - (R_{700})^{-1})/((R_{550})^{-1} + (R_{700})^{-1} \quad (2)$$

where R is reflectance, and the subscript indicates the wavelength (nm). The equation of NARI is mathematically identical to the NDAI. However, the NARI requires narrow green (550 nm) and red edge (700 nm) reflectance that can be acquired from satellite imagery. In contrast, the NDAI uses a wide range of green and red wavebands in the index calculation. We did not include NARI in the evaluation of the performance of various indices, since our equipment does not measure red-edge reflectance and substituting red for red-edge makes NARI identical to NDAI.

To assess the performance of NDAI in the prediction of anthocyanin content, we compared it to previously-developed indices. Gitelson et al. (2001) suggested the anthocyanin reflectance index (ARI):

$$ARI = (R_{550})^{-1} - (R_{700})^{-1} \quad (3)$$

where R is reflectance, and the subscript indicates the wavelength (nm). The ARI described anthocyanin concentrations in maple (*Acer platanoides*), dogwood (*Cornus alba*), geranium (*Pelargonium zonale*), and cotoneaster (*Cotoneaster alaunica*) leaves, with a coefficient of determination (R^2) great than 0.8 in each species. To calculate the ARI, the pixel intensities in each image were divided by 255, to convert them into reflectance values. Since the TopView system does not use the same wavelengths as the ARI, we used the green and red image, taken under the peak at 516 and 664 nm, the closest spectral images to the wavebands suggested for ARI (Figure 1).

Steele et al. (2009) introduced the modified anthocyanin content index (mACI):

$$mACI = R_{940}/R_{530} \quad (4)$$

where R is reflectance, and the subscript indicates the wavelength. Due to the availability of spectral images from the TopView system, we calculated mACI using the NIR and green image taken under the peaks at 861 and 516 nm (Figure 1B).

Gitelson et al. (2006) presented the modified anthocyanin reflectance index (mARI):

$$mARI = ((R_{530-570})^{-1} - (R_{690-710})^{-1}) \times R_{\text{longer than } 760} \quad (5)$$

where R is reflectance and the subscript indicates the range of wavelengths. The ranges in Eq. 5 represent the green, red, and near infrared spectrum, respectively. Due to the limitations of the TopView system, images taken under the LEDs with peaks at 516, 664, and 861 nm were used to calculate mARI.

Gamon and Surfus (1999) proposed the Red : Green ratio index (RGI) for anthocyanin prediction:

$$RGI = (R_{600-699})/(R_{500-599}) \quad (6)$$

where R is reflectance and the subscript indicates the range of wavelengths. The RGI was calculated from the green and red images taken under the LEDs with peaks at 516 and 664 nm (Figure 1). The RGI was correlated with anthocyanin concentration (R^2 = 0.92) in *Quercus agrifolia*.

A hand-held anthocyanin content meter (ACM-200 plus; Opti-Science, Inc., Hudson, NH, USA) was also used to measure the anthocyanin content index (ACI). This meter measures absorptance in the green (peak at 530 nm) and NIR (peak at 931 nm) to estimate anthocyanin concentrations (van den Berg and Perkins, 2005), likely with similar results as the mACI.

2.4 Automated image analysis

We wrote a program in Python (v. 3.8) using the OpenCV library (v. 4.5.4) to process multispectral images and calculate indices for anthocyanin content (see Script S1). The program reads monochrome images taken under different LED spectrums and extracts the intensity of each pixel from those images (0 to 255). The pixel intensities were then used to calculate anthocyanin content indices directly or after converting the pixel intensities into reflectance values by dividing by 255. The program creates a two-dimensional matrix, containing x- and y- coordinates and the corresponding index value. The matrix was then visualized as an index image of plant objects after background removal using intensity-based thresholding, followed by despeckling. The program generates a histogram, an average, and a standard deviation of all index values from only plant objects. The standard deviation is a measure of anthocyanin variability within the plant objects.

2.5 A low-cost phenotyping system for anthocyanin prediction

To evaluate the feasibility of using the NDAI for predicting anthocyanin concentration in a cost-effective manner, a Linux-based microcomputer (Raspberry Pi 4 model B, Raspberry Pi Foundation, Cambridge, UK) and an RGB camera (Raspberry Pi Camera Module v2.1, Raspberry Pi Foundation, Cambridge, UK) were used. The low-cost and automated imaging systems were installed in the growth chambers during the temperature/exposure time study to acquire top-view images of the lettuce plants at each harvest. Color images were taken every 10 minutes through the 'Crontab' command in Terminal in Raspberry Pi OS, which enables scheduling of a periodic task such as running a Python script at a given interval. The Python script for the camera operation was from the website of the Raspberry Pi Foundation (projects.raspberrypi.org/en/projects/getting-started-with-picamera) (Figure 2A). No gray-scale, or other calibration was used for this camera, to keep the method as simple as possible. The light spectrum in the growth chambers, spectral responses of the RGB camera, and combination of the spectral sensitivity of the RGB camera and the light spectrum in the growth chambers are visualized in Figure 2. The combination of them estimates the spectral response of each color channel during image acquisition under this particular light spectrum. Peak sensitivities of the system in the blue, red, and green color channels were 450, 560, and 597 nm, respectively. The spectral response of each color channel was broad compared to the LED spectrum in the TopView system. The red, green, and blue channels all have at least some sensitivity across the entire 400 – 700 nm range (Figure 2).

The images were analyzed with a modified program of the multispectral image analysis program that reads RGB images and extracts the pixel intensity of each color channel (see Script S2). The plant segmentation was established based on the red to blue ratio that showed strong contrast in the pixel intensity between plant objects and the background. The pixel intensity of plants in the blue was lower than that in the red, due to the higher absorptance in the blue spectrum. The background (metallic bottom of the growth chamber) did not show such differences in pixel intensity in the blue and red channels. With intensity-based thresholding based on the red to blue ratio, the program could separate plant objects from background and binary images were created. These binary images were then processed using erosion and dilation to improve plant segmentation. The algorithm for this segmentation will likely need to be modified for images taken under different conditions. The pixel intensity values of the pixels representing plants and for each

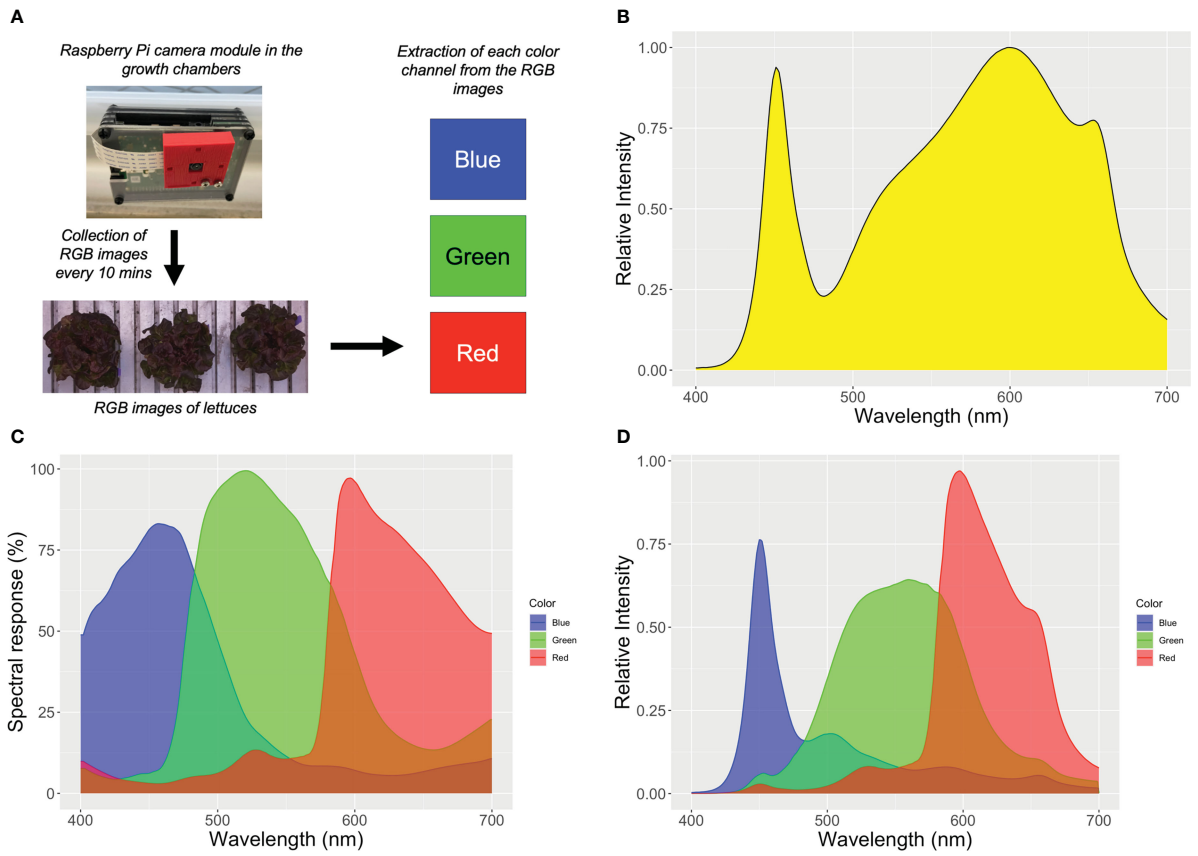


FIGURE 2
Schematic representation of acquisition process of green and red spectral images by the Raspberry Pi Camera Module V2.1 **(A)**, the normalized spectrum of the warm-white LED light fixtures in the growth chambers **(B)**, the spectral sensitivity of the camera module (Pagnutti et al., 2017) **(C)**, and the combination of the spectral sensitivity of the camera and the light spectrum in the growth chambers, showing the spectral sensitivity during acquisition of each spectral image **(D)**.

color channel were then used to calculate NDAI. The NDAI from the RGB images was compared to the anthocyanin concentrations of the top layer of the lettuce plants. The canopy NDAI from the RGB images also was compared with the canopy NDAI from multispectral images to assess the performance of the low-cost anthocyanin phenotyping system.

2.6 Statistical analyses

The performance of NDAI for anthocyanin prediction was evaluated by comparing it with other reflectance indices derived from multispectral images of leaf discs and plant canopy and the corresponding anthocyanin concentrations. Anthocyanin indices of leak discs, derived from equations 1 and 3 to 6, and ACI from the hand-held meter were compared with the measured anthocyanin concentrations. The coefficient of determination (R^2), root mean square error (RMSE), and Akaike information criterion (AIC) were calculated from the regression analyses between the indices and anthocyanin concentrations (R Core Team, 2022).

In general, an R^2 value higher than 0.7 is considered indicative of a good model that can explain significant amount of variance, while a higher value indicates a better model fit (Frost, 2019). A lower RMSE or AIC value indicates a better model fit, but the absolute value of RMSE or AIC is not informative in determining whether a particular value indicates a good model fit (Baguley, 2018). Following the criteria of a better model in these statistical metrics, we selected the best performing model for prediction of anthocyanin concentration in the leaf discs.

Using the best-performing index model, we also tested its feasibility in prediction of anthocyanin concentration of a canopy layer using the multispectral images acquired by the TopView system and RGB images taken by the Raspberry Pi-RGB camera imaging system. To evaluate the correlation between the canopy anthocyanin concentration and the index values derived from each imaging system, we calculated R^2 and RMSE.

3 Results

3.1 Evaluation of anthocyanin content indices

Anthocyanin concentrations of the 50 leaf discs ranged from 108 to 1673 μg g^{-1} of fresh weight. Higher anthocyanin concentrations of the leaf discs coincided with higher values of all image-derived indices, and can be observed within the false color images representing each index (Figure 3). The anthocyanin concentrations in these example images differed by a factor of 3.94 and the range of values obtained from the various indices differed greatly. Although the average RGI of the high-anthocyanin leaf discs was higher than that of the low-anthocyanin leaf discs, the difference was relatively small (1.38×). For the other indices, the index values for the low- and high-anthocyanin leaf discs differed by 2.05 to 5.00× (Figure 3).

These index images show that the distribution of anthocyanins within leaf discs was not uniform (Figure 3), although the leaf discs were small (3.1 cm^2). Especially, the NDAI, ARI, and mARI images were able to display non-uniformity that was not visible in the corresponding RGB image. On the other hand, the mACI and RGI image had a somewhat limited depiction of the heterogeneous anthocyanin distribution in the leaf discs, resulting in a low coefficient of variation. The coefficient of variation in the mACI and RGI images was 0.33 and 0.15 for the higher anthocyanin concentration leaf disc and 0.13 and 0.21 for the lower anthocyanin concentration leaf disc, respectively. Images of the other indices, had coefficients of variation of 0.57 to 0.82 and 1.94 to 2.58, in the higher and low anthocyanin concentration, respectively.

An increase in anthocyanin concentration was associated with a linear or quadratic increase in all anthocyanin indices (Figure 4). As anthocyanin concentrations of the leaf discs increased from 108 to 1673 μg g^{-1} of FW, the average NDAI, ARI, mACI, mARI, and RGI ranged from -0.062 to 0.192, -0.950 to 3.680, 1.600 to 6.786, -0.508 to 1.909, and 0.900 to 1.496, respectively.

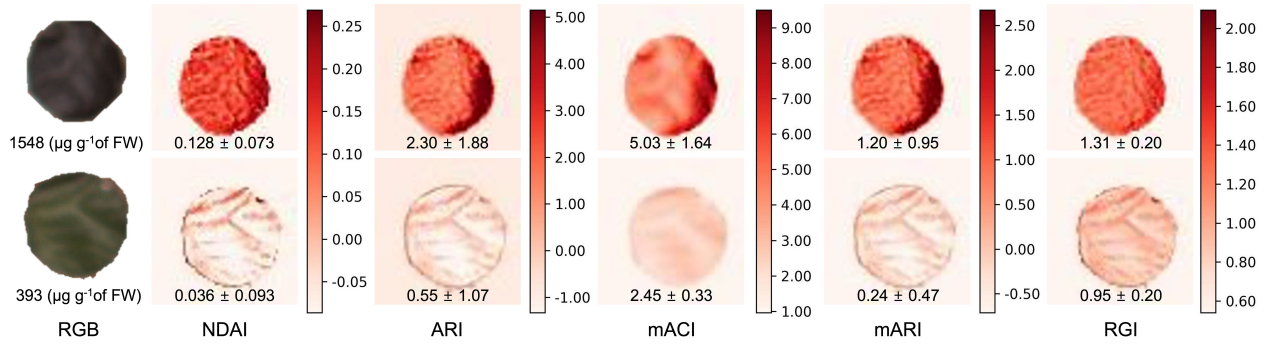


FIGURE 3
Images of various indices of anthocyanin concentration based on multispectral images. The leaf disc at the top had a higher anthocyanin concentration than the bottom one. From left to right: RGB (color) image, normalized difference anthocyanin index (NDAI; Eq. 1), anthocyanin reflectance index (ARI; Eq. 3), modified anthocyanin content index (mACI; Eq. 4), modified anthocyanin reflectance index (mARI; Eq. 5), and red to green ratio index (RGI; Eq. 6). The values in each image represents the average ± standard deviation of the corresponding index value, while the RGB images show the anthocyanin concentration of the leaf discs.

The NDAI, ARI, mARI, and RGI (Figure 4 and Table 1) had higher R^2 than other indices (0.80< R^2< 0.89). These indices also had lower RMSE and AIC values (0.02< RMSE< 0.52, -237.6< AIC< 82.1). The NDAI had the highest R^2 (0.89), lowest RMSE (0.02) and lowest AIC (-237.6) among the indices for predicting anthocyanin concentration. At the same time, the RGI had R^2 of 0.89, RMSE of 0.05, and AIC of -149.7. The ACI and mACI resulted in lower values of these evaluation metrics (Figure 4 and Table 1), resulting in R^2 values of 0.36 and 0.34, RMSEs of 2.04 and 0.94, and AICs of 217.4 and 140.8 (Table 1), respectively. Changing the units of anthocyanin concentration to an area-based unit (mg m^{-2}) did not affect the trends in the statistical metrics.

Differences in the statistical metrics can be used to determine which wavelengths to use for an anthocyanin index. The NDAI, ARI, and RGI all use reflectance in the green and red, while mARI also uses NIR, whereas ACI and mACI use reflectance in the green and NIR. This suggests that the use of the red waveband is

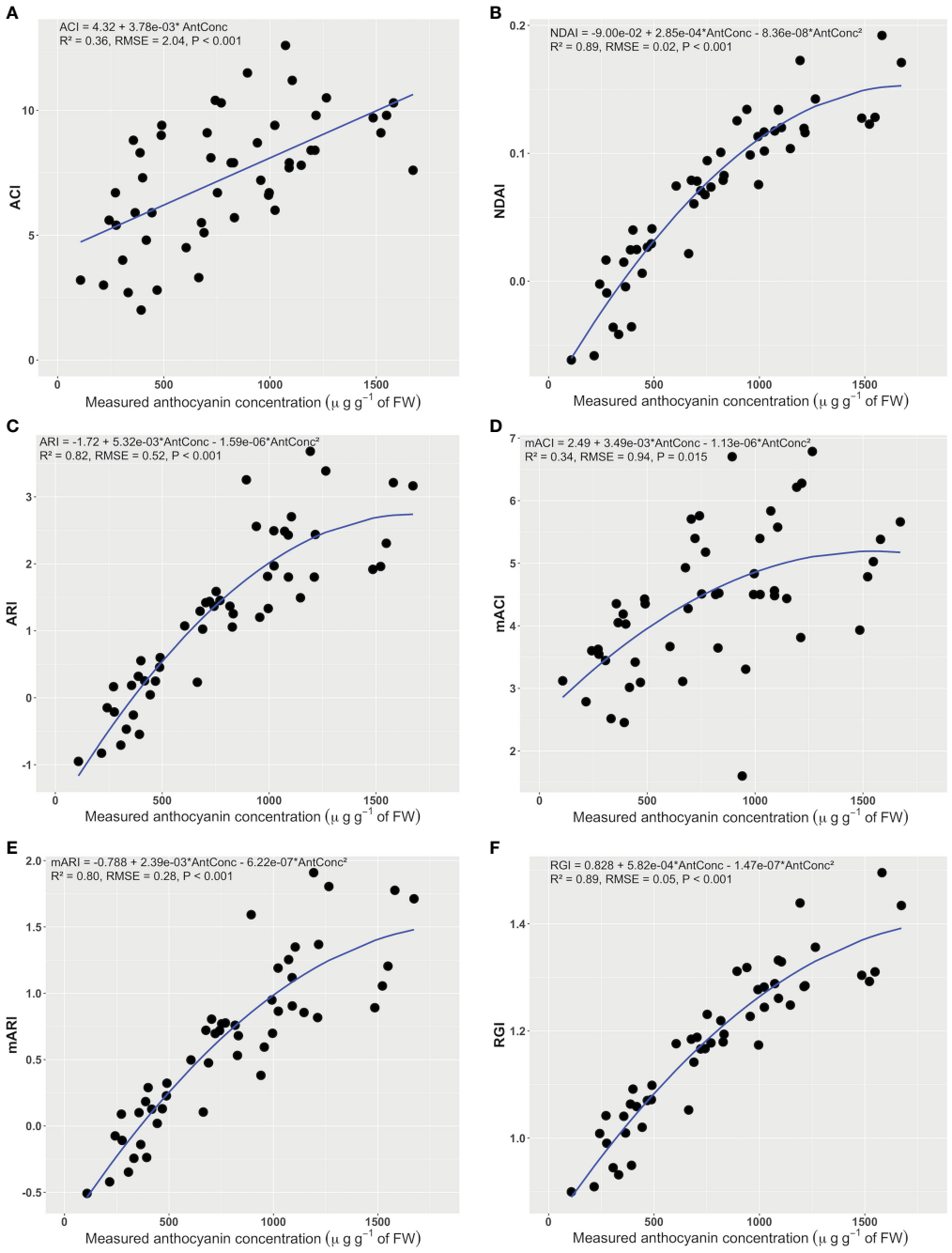


FIGURE 4
The relationship between the anthocyanin concentration of leaf discs and the corresponding anthocyanin content index of a hand-held anthocyanin content meter (ACI) **(A)**, anthocyanin content index **(B)**, anthocyanin reflectance index (ARI) **(C)**, modified anthocyanin reflectance index (mARI) **(D)**, modified anthocyanin content index (mACI) **(E)**, and red to green ratio index (RGI) **(F)** (n = 50). The regression summaries include the regression equation, coefficient of determination, root mean square error, and p-value. The blue curves show the regression equations.

TABLE 1 Statistical summaries of various indices for anthocyanin content (n = 50).

Index	R^2	RMSE	AIC	*p*-Value
ACI	0.36	2.04	217.4	< 0.001
NDAI	0.89	0.02	-237.6	< 0.001
ARI	0.82	0.52	82.1	< 0.001
mACI	0.34	0.94	140.8	0.015
mARI	0.80	0.28	19.6	< 0.001
RGI	0.89	0.05	-149.7	< 0.001

ACI, Anthocyanin content index measured by hand-held anthocyanin content meter; NDAI, normalized difference anthocyanin index; ARI, anthocyanin reflectance index; mACI, modified anthocyanin content index; mARI, modified anthocyanin reflectance index; and RGI, red and green ratio index.

preferable to the use of NIR, likely because the red reflectance can help account for different chlorophyll concentrations.

The NDAI predicted anthocyanin concentration slightly better than the RGI, based on its lower AIC value and RMSE (Table 1). Based on these results, the best index for image-based anthocyanin phenotyping is NDAI, while RGI also performs well. Note that all indices, except for ACI, had a non-linear relationship with the measured anthocyanin concentration. Interestingly, the commercially-available anthocyanin meter, which measures ACI, performed poorly (R^2 = 0.36).

3.2 Canopy NDAI imaging

Canopy NDAI images were obtained by using the pixel intensities from canopy images taken under red and green light (Figure 5). Canopy NDAI images of 'Rouxai' and 'Teodore' clearly showed the difference in their anthocyanin concentrations (Figure 5C). The canopy ACI images provide spatial information by depicting the distribution of anthocyanins within the canopy.

The anthocyanin concentrations of top canopy layer the red lettuce cultivars ranged from 251 to 928 μg g^{-1} of FW, while the average canopy NDAI increased from -0.062 to 0.211 with increases in the anthocyanin concentrations (Figure 6). Anthocyanin concentrations in the top canopy layer had a positive correlation (R^2 of 0.73 and RMSE = 0.04) with the average canopy NDAI. This relationship was not a strong as that for the leaf discs, likely because of the non-uniform distribution of the anthocyanins. Anthocyanins were extracted from the top layer of the canopy, but it was not possible to sample the exact same part of the canopy that was visible in the images. However, the high R^2 and low RMSE of the canopy NDAI versus extracted anthocyanin concentration indicate that this method is useful to predict anthocyanin concentrations using canopy multispectral images.

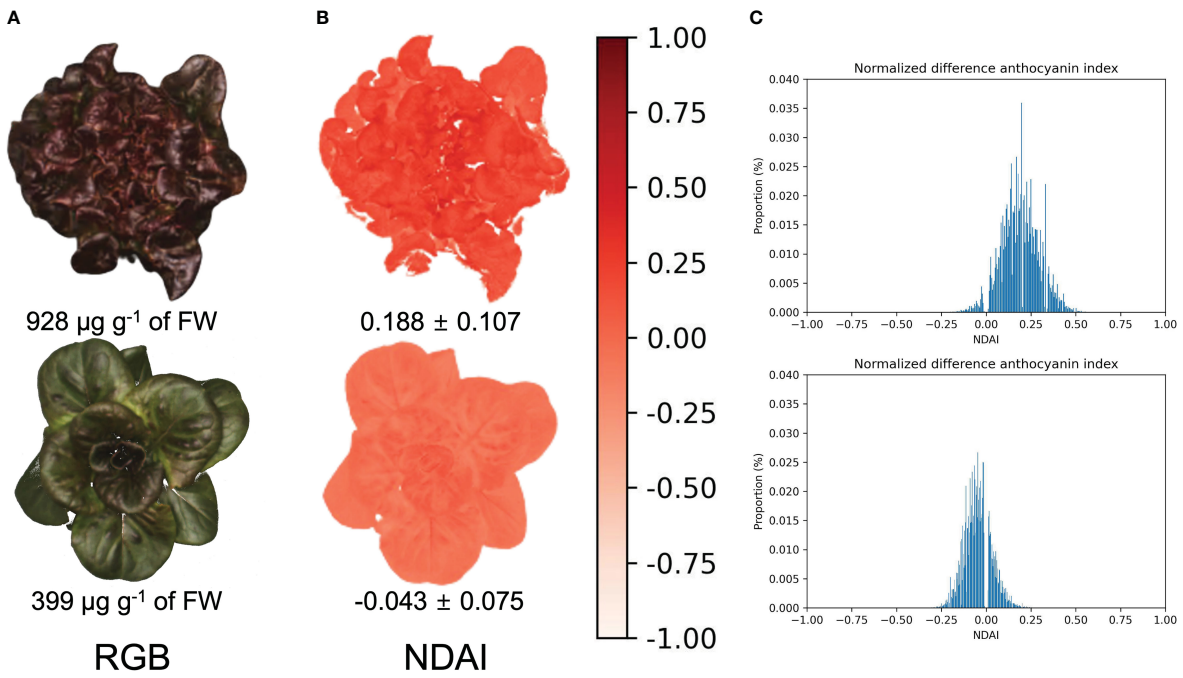


FIGURE 5
Canopy RGB **(A)** and normalized difference anthocyanin index (**B**; NDAI) images of red lettuce cultivars 'Rouxai' (top) and 'Teodore' (bottom) and their corresponding histograms **(C)**. The scale bar on the right side represents the NDAI values within these NDAI images. Values below the NDAI images represent mean ± standard deviation of NDAI within the NDAI images, while values below the RGB images represent the anthocyanin concentration.

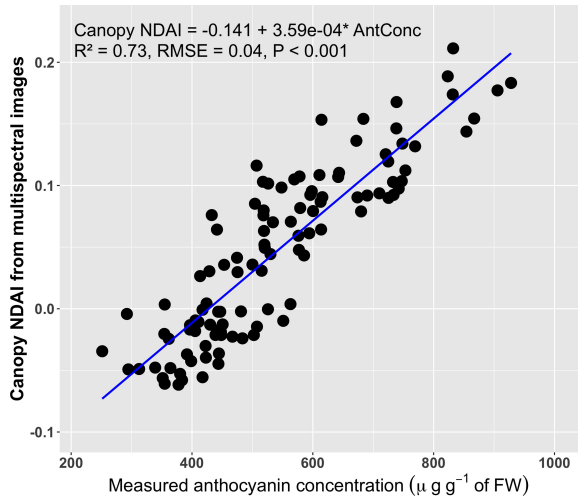


FIGURE 6
Correlation between the canopy normalized difference anthocyanin index (NDAI, averaged over the entire part of the canopy visible in the images) derived from multispectral images and the anthocyanin concentrations of the top layer of the canopy (n = 108). The blue line is the regression line.

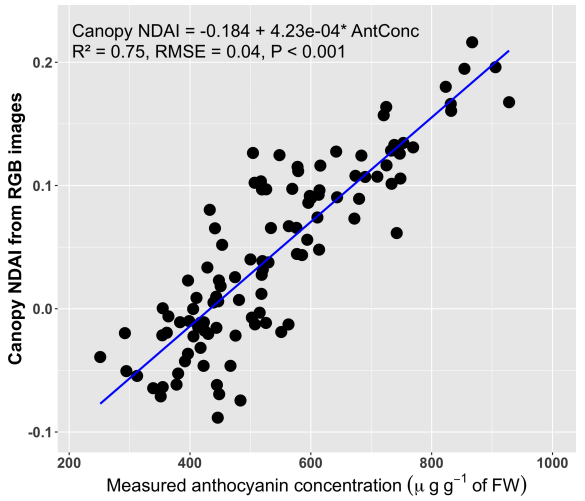


FIGURE 7
Correlation between the normalized difference anthocyanin index (NDAI) derived from RGB images and averaged over the entire visible part of the canopy and the canopy anthocyanin concentration (n = 108) **(A)**. The blue line is the regression line.

3.3 A low-cost imaging system for canopy NDAI

The red and green pixels of the RGB camera under the white LED fixtures in the growth chambers were used as proxies for the green and red images from the TopView multispectral imaging system. The green and red channels of the TopView system had peaks at wavelengths of 516 and 664 nm, respectively (Figure 1). The red channel of the RGB camera had a sensitivity of >50% at wavelengths from 580 to 700 nm with a peak at 596 nm, but was also sensitive to wavelengths below 580 nm (Figure 2B). The green channel had a sensitivity above 50% at wavelengths from 477 to 595 nm with a peak at 521 nm (Figure 2B). Because of the different spectral responses of the RGB camera and multi-spectral imaging system, we tested whether the canopy NDAI from the multispectral images and RGB camera are correlated. Given the different spectral response of the two systems, we did not expect the values to be identical.

The average canopy NDAI based on the RGB images was positively correlated with the anthocyanin concentration (R^2 = 0.75, RMSE = 0.04; Figure 7). The average canopy NDAI from the RGB images ranged from -0.088 to 0.216, similar to the range of the canopy NDAI from the multispectral images, which ranged from -0.061 to 0.211.

Likewise, the canopy NDAI from the multispectral images and from the RGB images were positively correlated (R^2 = 0.87, RMSE = 0.03; Figure 8). Comparing the regression model (black line) to the 1:1 line (blue line) suggests that the NDAI from RGB images was slightly overestimated at NDAI values< 0.05 and underestimated at NDAI > 0.05. Such differences might be associated with errors in background removal during the RGB image processing, differences in the light spectrum during the RGB image acquisition, angle of the camera, and spectral sensitivity of the two methods. Despite these differences between the two imaging approaches, there was a strong correlation between the two methods, indicating that the RGB images can be used to estimate NDAI.

4 Discussion

Evaluation of the image-derived anthocyanin indices demonstrates the feasibility of image-based phenotyping for estimating anthocyanin concentrations in lettuce. Our initial testing was conducted using leaf discs with the multispectral

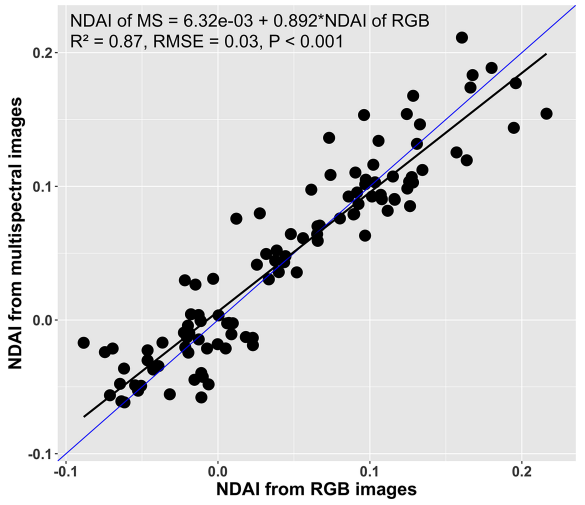


FIGURE 8
Correlation between canopy normalized difference anthocyanin index (NDAI) derived from multispectral and RGB images (n = 108). The black line is the regression line, while the blue line is a 1:1 relation line.

imaging system. Because using a narrow area of leaf and a reliable spectral imaging system can be considered a reference condition for development of a precise prediction model for anthocyanin concentration. The selected NDAI, therefore, resulted in R^2 of 0.73 for the correlation between the canopy anthocyanin concentration and the canopy NDAI using the multispectral imaging system. Furthermore, the canopy NDAI derived from the RGB images resulted in R^2 of 0.75, demonstrating that a potential method of whole plant or crop anthocyanin phenotyping at a low cost.

In addition, NDAI values obtained from the multispectral imaging system and RGB images were similar, indicating that a regular digital camera, combined with a broad-band white light source may be all that is required for NDAI imaging. All tested anthocyanin indices showed a positive relationship with leaf disc anthocyanin concentrations, but the performance of the different indices varied greatly (Figure 4 and Table 1). Although our multispectral images (Figure 1) do not use the same wavebands as those typically used for ARI, mARI, and mACI (Eq. 3-5) (e.g. ARI and mARI required 700 nm while we used 664 nm and mACI required 940 nm while we used 861 nm) (Gitelson et al., 2006; Gitelson et al., 2009; Steele et al., 2009), the image-derived indices that used the green and red part of the spectrum (NDAI, ARI, mARI, and RGI) achieved a reasonably good model fit in prediction of anthocyanin concentrations ($R^2 \geq 0.8$, RMSE ≤ 0.52, AIC ≤ 82.1).

That success is associated with the role of the green and red spectrum in the anthocyanin indices. The variation in reflectance in the green part of the spectrum (510 to 550 nm) is partly determined by anthocyanin concentration, because anthocyanins absorb green photons effectively (Neill and Gould, 2000; Gitelson et al., 2001). However, chlorophyll also absorbs green light. Reflectance in the red part of the spectrum (660 nm to 710) can account for variability in chlorophyll concentration due to the relationship between chlorophyll concentration and absorptance in this part of the spectrum (Lichtenthaler, 1987; Gitelson et al., 1996). The NDAI, ARI, mARI, and RGI were able to quantify anthocyanin concentration accurately by removing interference from variable chlorophyll concentrations (Gitelson et al., 2009).

Not all indices attempt to correct for the influence of chlorophyll. The ACI and mACI, use the green and NIR part of the spectrum, resulting in R^2 values of 0.36 and 0.34, respectively. This was consistent with the report by Steele et al. (2009), who reported a weak relationship between mACI and anthocyanin content of grape vine leaves (R^2 = 0.06). Slaton et al. (2001) reported that the reflectance in the NIR is associated with characteristics of leaf structure, such as leaf thickness, rather than chlorophyll content. Therefore, using NIR instead of red has no benefits in accounting for chlorophyll. The mARI uses NIR as a third waveband to account for variability in leaf thickness and light scattering within the leaf (Gitelson et al., 2006), but this did not have any distinctive advantages over the indices using only green and red, such as NDAI, ARI and RGI based on statistical metrics.

The form of the index equation determines the range of values the index can have. NDAI, adopting the NDVI equation, is constrained to the range between -1 and 1, but in reality ranged only from -0.25 to 0.5 (Figure 5). The other indices do not have constrained ranges for their value. This could potentially result in extreme values, making averaging of multiple readings potentially meaningless. Since the NDAI had slightly better statistical performance than RGI and the other indices (Table 1) and is constrained between -1 and 1, we conclude that NDAI has advantages over the other indices.

There are many successful examples of anthocyanin predictions based on hyperspectral imaging combined with machine learning techniques (Chen et al., 2015; Askey et al., 2019; Simko, 2020; Cho et al., 2021; Kim et al., 2021). However, these machine learning models require hyperspectral imaging systems with similar wavelengths as used to develop the machine learning model. Thus, such models may not perform well when applied to data collected with different hyperspectral imaging systems or under different lighting conditions. In addition, hyperspectral imaging systems provide much more information than our NDAI calculation requires, and are generally expensive. The high cost of technical requirements of hyperspectral imaging and machine learning makes widespread adoption difficult.

NDAI requires only red and green images, which can be acquired from RGB images or simple multispectral imaging systems. The NARI, having the identical equation as NDAI, successfully detected variability in anthocyanin concentrations in mountain shrublands (Bayle et al., 2019). However, it uses satellite imagery as its source of spectral reflectance, and relies on narrow green and red-edge wavebands. Furthermore, the NARI has not been evaluated by comparing index values to measured anthocyanin concentrations. On the other hand, the NDAI, with the low-cost RGB imaging or the multispectral imaging, had a good correlation with canopy anthocyanin concentrations. That suggests that pictures taken by a cellular phone or any color camera, which can capture red and green color channels, can be used to calculate NDAI. Due to its simplicity and the lack of need for a narrow specific waveband, RGB-derived NDAI can be easily implemented at the single leaf or whole canopy scale, including in remote sensing and indoor horticultural production. Such a system can be built using an RGB camera and Raspberry Pi microcomputer, at a cost of about $60. The simplicity, low cost, and automated processing can make multispectral imaging available to a wide range of researchers and growers, who need such technology but cannot afford expensive systems.

Some limitations of this image-based phenotyping approach are associated with imaging of plants from above. The images largely capture the properties of the cell layers near the top of the canopy. If anthocyanins are present on the bottom side of leaves, imaging may not detect them. In addition, the camera can only see one layer of leaves and the system thus provides no information regarding lower leaves if they are obscured by the upper leaves. Segmentation of plant objects is also a prerequisite to obtain anthocyanin indices. Finally, NDAI values, as well as other indices, are affected by the spectrum of the light under which the images are taken. As long as the light spectrum is the same for all images, calculated NDAI values will still reflect differences in anthocyanin content. However, if the light spectrum changes, this poses a challenge for comparing NDAI values. This might be solved by *in situ* spectral calibration procedures using an in-scene reference card (Davies et al., 2022).

The monitoring of the anthocyanin content using NDAI has a wide range of agricultural and ecological applications. Not surprisingly, the optical properties of anthocyanins in fruit and vegetative tissues are similar. Therefore, NDAI may be used to assess ripeness of fruits that accumulate anthocyanins during ripening, such as grapes and strawberries. Indeed, Underhill et al. (2020) showed that the RGB color space differentiated variability in grape skin color, associated with different anthocyanin concentrations. Strawberries with higher concentrations of anthocyanin had lower pixel intensities in the green spectrum (the lowest intensity was near 530 nm) based on hyperspectral imaging (Cho et al., 2021). Monitoring anthocyanins in fruits and vegetables will be beneficial not only in assessing phenotypic variation in anthocyanins in a non-destructive manner, but also in automating post-harvest quality evaluation or for robotic harvesting. Finally, the integration of NDAI phenotyping and environmental control systems used in controlled environment agriculture may be used for dynamic environmental control to stimulate anthocyanin production in many vegetables and fruits.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

CK and MI conceived the original idea; CK and MI designed the experiment; and CK collected and analyzed the data and wrote the first draft, with inputs from MI. Both authors contributed to the article and approved the submitted version.

Funding

This work was funded by USDA-NIFA-SCRI (Award Number 2018-51181-28365), Project 'Lighting Approaches to Maximize Profits'.

Acknowledgments

We thank Drs. Savithri Nambeesan and Yi-Wen Wang for their assistance with anthocyanin extraction.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1155722/full#supplementary-material

DOCUMENT S1
Manual for image analysis script.

SCRIPT S1
Python script for normalized difference anthocyanin index using multispectral images.

SCRIPT S2
Python script for normalized difference anthocyanin index using RGB images.

References

Askey, B. C., Dai, R., Lee, W. S., and Kim, J. (2019). A noninvasive, machine learning–based method for monitoring anthocyanin accumulation in plants using digital color imaging. *Appl. Plant Sci.* 7 (11), e11301. doi: 10.1002/aps3.11301

Baguley, T. (2018). *Serious stat: A guide to advanced statistics for the behavioral sciences* (London, United Kingdom: Bloomsbury Publishing).

Bayle, A., Carlson, B., Thierion, V., Isenmann, M., and Choler, P. (2019). Improved mapping of mountain shrublands using the sentinel-2 red-edge band. *Remote Sens.* 11, 2807. doi: 10.3390/rs11232807

Boldt, J. K., Meyer, M. H., and Erwin, J. E. (2014). Foliar anthocyanins: A horticultural review. *Horticultural Reviews* 42, 209–251. doi: 10.1002/9781118916827.ch04

Chen, S., Zhang, F., Ning, J., Liu, X., Zhang, Z., and Yang, S. (2015). Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. *Food Chem.* 172, 788–793. doi: 10.1016/j.foodchem.2014.09.119

Cho, J.-S., Lim, J. H., Park, K. J., Choi, J. H., and Ok, G. S. (2021). Prediction of pelargonidin-3-glucoside in strawberries according to the postharvest distribution period of two ripening stages using VIS-NIR and SWIR hyperspectral imaging technology. *LWT* 141, 110875. doi: 10.1016/j.lwt.2021.110875

Costa, C., Schurr, U., Loreto, F., Menesatti, P., and Carpentier, S. (2019). Plant phenotyping research trends, a science mapping approach. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01933

Das Choudhury, S., Samal, A., and Awada, T. (2019). Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00508

Davies, M., Stuart, M. B., Hobbs, M. J., Mcgonigle, A. J. S., and Willmott, J. R. (2022). Image correction and *In situ* spectral calibration for low-cost, smartphone hyperspectral imaging. *Remote Sens.* 14, 1152. doi: 10.3390/rs14051152

Del Valle, J. C., Gallardo-Lopez, A., Buide, M. L., Whittall, J. B., and Narbona, E. (2018). Digital photography provides a fast, reliable, and noninvasive method to estimate anthocyanin pigment concentration in reproductive and vegetative plant tissues. *Ecol. Evol.* 8, 3064–3076. doi: 10.1002/ece3.3804

Frost, J. (2019). *Regression analysis: An intuitive guide for using and interpreting linear models* (State College, PA, United States: Statisics By Jim Publishing).

Gamon, J. A., Serrano, L., and Surfus, J. S. (1997). The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia* 112, 492–501. doi: 10.1007/s004420050337

Gamon, J. A., and Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* 143, 105–117. doi: 10.1046/j.1469-8137.1999.00424.x

Gitelson, A. A., Chivkunova, O. B., and Merzlyak, M. N. (2009). Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *Am. J. Bot.* 96, 1861–1868. doi: 10.3732/ajb.0800395

Gitelson, A. A., Keydan, G. P., and Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Res. Lett.* 33, LL11402. doi: 10.1029/2006GL026457

Gitelson, A. A., Merzlyak, M. N., and Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochem. Photobiol.* 74, 38–45. doi: 10.1562/0031-8655(2001)074<0038:opaneo>2.0.co;2

Gitelson, A. A., Merzlyak, M. N., and Lichtenthaler, H. K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *J. Plant Physiol.* 148, 501–508. doi: 10.1016/S0176-1617(96)80285-9

Khoo, H. E., Azlan, A., Tang, S. T., and Lim, S. M. (2017). Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr. Res.* 61, 1361779. doi: 10.1080/16546628.2017.1361779

Kim, H., Yoo, J. H., Park, S. H., Kim, J.-S., Chung, Y., Kim, J. H., et al. (2021). Measurement of environmentally influenced variations in anthocyanin accumulations in brassica rapa subsp. chinensis (Bok choy) using hyperspectral imaging. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.693854

Lee, J., Durst, R. W., and Wrolstad, R. E. (2005). Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method: Collaborative study. *J. AOAC Int.* 88, 1269–1278. doi: 10.1093/jaoac/88.5.1269

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods Enzymology* 148, 350–382. doi: 10.1016/0076-6879(87)48036-1

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in solanaceous vegetables: A review. *Front. Chem.* 6. doi: 10.3389/fchem.2018.00052

Maes, W. H., and Steppe, K. (2019). Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* 24, 152–164. doi: 10.1016/j.tplants.2018.11.007

Merzlyak, M. N., Chivkunova, O. B., Solovchenko, A. E., and Naqvi, K. R. (2008). Light absorption by anthocyanins in juvenile, stressed, and senescing leaves. *J. Exp. Bot.* 59, 3903–3911. doi: 10.1093/jxb/ern230

Neill, S., and Gould, K. S. (2000). Optical properties of leaves in relation to anthocyanin concentration and distribution. *Can. J. Bot.* 77, 1777–1782. doi: 10.1139/b99-153

Pagnutti, M., Ryan, R., Cazenavette, G., Gold, M., Harlan, R., Leggett, E., et al. (2017). Laying the foundation to use raspberry pi 3 V2 camera module imagery for scientific and engineering purposes. *J. Electronic Imaging* 26, 013014. doi: 10.1117/1.JEI.26.1.013014

R Core Team (2022). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).

Rouse, J. R. J., Haas, R. H., Deering, D., Schell, J., and Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*. NASA Technical Reports. Available at: https://ntrs.nasa.gov/citations/19730017588.

Simko, I. (2020). Predictive modeling of a leaf conceptual midpoint quasi-color (CMQ) using an artificial neural network. *Sensors* 20, 3938. doi: 10.3390/s20143938

Slaton, M. R., Raymond Hunt, E., and Smith, W. K. (2001). Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am. J. Bot.* 88, 278–284. doi: 10.2307/2657019

Steele, M. R., Gitelson, A. A., Rundquist, D. C., and Merzlyak, M. N. (2009). Nondestructive estimation of anthocyanin content in grapevine leaves. *Am. J. Enology Viticulture* 60, 87–92. doi: 10.5344/ajev.2009.60.1.87

Underhill, A. N., Hirsch, C. D., and Clark, M. D. (2020). Evaluating and mapping grape color using image-based phenotyping. *Plant Phenomics* 2020, 1–11. doi: 10.34133/2020/8086309

van den Berg, A. K., and Perkins, T. D. (2005). Nondestructive estimation of anthocyanin content in autumn sugar maple leaves. *HortScience* 40, 685–686. doi: 10.21273/HORTSCI.40.3.685

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 April 2023
DOI 10.3389/fpls.2023.1150958

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Xi Qiao,
Agricultural Genomics Institute at Shenzhen (CAAS), China
Lingxian Zhang,
China Agricultural University, China

*CORRESPONDENCE
Guoxu Liu
liuguoxu@wfu.edu.cn
Suk Chan Kim
sckim@pusan.ac.kr

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 25 January 2023
ACCEPTED 20 March 2023
PUBLISHED 03 April 2023

CITATION
Mbouembe PLT, Liu G, Sikati J, Kim SC and Kim JH (2023) An efficient tomato-detection method based on improved YOLOv4-tiny model in complex environment.
Front. Plant Sci. 14:1150958.
doi: 10.3389/fpls.2023.1150958

COPYRIGHT
© 2023 Mbouembe, Liu, Sikati, Kim and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An efficient tomato-detection method based on improved YOLOv4-tiny model in complex environment

Philippe Lyonel Touko Mbouembe[1], Guoxu Liu[2*], Jordane Sikati[1], Suk Chan Kim[1*] and Jae Ho Kim[1]

[1]Department of Electronics Engineering, Pusan National University, Busan, Republic of Korea, [2]School of Computer Engineering, Weifang University, Weifang, China

Automatic and accurate detection of fruit in greenhouse is challenging due to complicated environment conditions. Leaves or branches occlusion, illumination variation, overlap and cluster between fruits make the fruit detection accuracy to decrease. To address this issue, an accurate and robust fruit-detection algorithm was proposed for tomato detection based on an improved YOLOv4-tiny model. First, an improved backbone network was used to enhance feature extraction and reduce overall computational complexity. To obtain the improved backbone network, the BottleneckCSP modules of the original YOLOv4-tiny backbone were replaced by a Bottleneck module and a reduced version of BottleneckCSP module. Then, a tiny version of CSP-Spatial Pyramid Pooling (CSP-SPP) was attached to the new backbone network to improve the receptive field. Finally, a Content Aware Reassembly of Features (CARAFE) module was used in the neck instead of the traditional up-sampling operator to obtain a better feature map with high resolution. These modifications improved the original YOLOv4-tiny and helped the new model to be more efficient and accurate. The experimental results showed that the precision, recall, F_1 score, and the mean average precision (mAP) with Intersection over Union (IoU) of 0.5 to 0.95 were 96.3%, 95%, 95.6%, and 82.8% for the improved YOLOv4-tiny model, respectively. The detection time was 1.9 ms per image. The overall detection performance of the improved YOLOv4-tiny was better than that of state-of-the-art detection methods and met the requirements of tomato detection in real time.

KEYWORDS

YOLOv4-tiny model, tomato detection, deep learning, computer vision, agriculture

1 Introduction

Recent advances of artificial intelligence technology have allowed wide applications in every area of life, including agriculture. For a decade, fruit detection has been a very active research direction. Detecting and sorting single crops plants such as oranges, apples, tomatoes, etc. are difficult and time intensive due to the number of varieties of the same

fruit and environment conditions such as cluster. With the development of artificial intelligence, this can be done by robots. Moreover, Computer vision and related algorithms have been applied to improve the efficiency, intelligence, and remote interactions of robots in complex agricultural environments (Cheng et al., 2020).

A series of traditional fruit detection and recognition algorithms have been proposed. Most of them used non-pattern methods such as color, texture, and geometry methods for fruit detection. Linker et al. (2012) used color and texture to classify green apples. However, sunlight and color-saturation variation which constitute the illumination variation had a large impact on their results. Furthermore, Zhao et al. (2016) developed a method based on segmented mature tomatoes from background using an optimal threshold on fusion image features, but illumination also affected their results. Moreover, an optimal threshold extracted from the intensity histogram of a red-color-difference enhanced image for apple recognition. But this method was restricted to ripe apples which present different color to the background. Liu et al. (2019) proposed a coarse-to-fine method for ripe tomato detection in greenhouse. A naive Bayes classifier combined with a histogram of oriented gradients was applied to recognize tomatoes. The method also used a colour analysis to remove false detection. But due to the low-level abstraction capabilities of hand-crafted features, it was difficult to adapt this method on complex environment changes. Finally, a shape analysis method for mature apple localization proposed by Kelman and Linker (2014) used a canny filter to find the edges in the image. The method also used a pre-processing operation and convexity test to detect the edges that correspond to three-dimensional convex objects. The performance was influenced by illumination and leaves that have similar convex surfaces to apples.

Since the traditional methods were based on handcrafted features, they had several drawbacks, such as low-level of feature extraction in certain conditions. These problems were conquered with the introduction of deep learning (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014). Deep learning techniques have great performance in many fields, including vision tasks (Kamilaris and Prenafeta-Boldu, 2018). First, Sa et al. (2016) merged multi-modal color (RGB) and near-infrared (NIR) information based on a Faster R-CNN (Ren et al., 2015) detector for fruit detection. The model obtained better result than previous models. However, it was difficult to detect small fruits, and the speed still needed to be improved for real-time detection for a harvesting robot. Rahnemoofar and Sheppard (2017) proposed a modified Inception-Resnet architecture (Szegedy et al., 2017) for fruit counting and achieved good results. However, the model just counted fruit and did not detect them. Furthermore, Mu et al. (2020) and Gao et al. (2020) proposed an R-CNN algorithm (Girshick et al., 2014) using ResNet (Szegedy et al., 2017) as a backbone network for the detection, counting and size estimation of green tomatoes. Afonso et al. (2020) used Mask R-CNN (Kaiming et al., 2017) to tomato datasets for detection. Many neural networks were used as backbone to extract feature map.

"You Only Look Once" (YOLO) models were proposed by Redmon et al. (2016); Redmon and Farhadi (2017); Redmon and Farhadi (2018); Bochkovskiy et al. (2020), and Wang et al. (2022) for object detection. They had great improvement in both speed and accuracy compared with the previous region proposal-based detectors (Girshick et al., 2014; Ren et al., 2015; Kaiming et al., 2017), which performed detection in a two-stage pipeline. YOLO models directly predicted the bounding boxes and their corresponding classes with single feed-forward network. There are some studies on fruit detection using YOLO models. Liu et al. (2020) developed a robust model on tomato detection named YOLO-Tomato based on YOLOv3. The traditional rectangular box was replaced with a circular bounding box to match the tomato target. The model achieved an AP of 96.40% with a detection time of 54 ms. Moreover, Fu et al. (2020) developed an algorithm based on improved YOLOv3-tiny to detect kiwifruits in orchard. Xu et al. (2020) proposed a fast method of detecting tomatoes in a complex scene for picking robots, and their experimental results showed that the F1 score was 91.92% with an inferential of 40.35 ms. Furthermore, Wang X et al. (2021) also proposed an algorithm based on YOLOv3-tiny to detect diseases of occlusion and overlapping tomato leaves. A YOLOv3-tiny-IRB algorithm was used to reduce layer-by-layer loss if information during network transmission. The model got a mAP of 93.1%. Furthermore, Arunabha and Jayabrata (2021) proposed a detection method for fine grain based on a modification of YOLOv4 model and achieve good results. In their model, the Dense Net (Gao et al., 2018) architecture was inserted in the backbone to enhance feature map extraction. The result showed the mAP was 96.29% with a detection time of 70.19 FPS. Rupareliya et al. (2022) proposed a deep learning-based tomato detection, where different versions of the YOLO architectures were used. Chenglin et al. (2022) proposed YOLO-PEFL model to detect pear flowers in the natural environment based on improved YOLOv4. The AP of the model was 96.71%, the model size was reduced by 80% approximatively, and the detection time was 0.027 s. Finally, Tang et al. (2023) proposed YOLO- Oleifera based on improved YOLOv4-tiny model and binocular stereo vision, and it achieved an AP of 92.07% with an average of 31 ms to detect each fruit image.

Although much research has been conducted on fruit detection in complex environment, the detection accuracy and the efficiency still need to be improved to meet the requirement of fruit detection under complicated conditions.

To address the above issues, an efficient tomato-detection method was proposed based on an improved YOLOv4-tiny model in this study. Figure 1 shows an overview of the improved model. The main contributions of this study are as follow:

1 A modified BottleneckCSP module was designed and inserted in the backbone network to enhance the feature extraction and to reduce the computational complexity,

2 A tiny version of CSP-SPP module was also designed and attached to the new backbone network to improve the receptive field,

3 The CARAFE module was used in the neck to get feature map with higher feature map,

4 Extensive experiments were conducted on the tomato datasets to show that the improved YOLOv4-tiny model

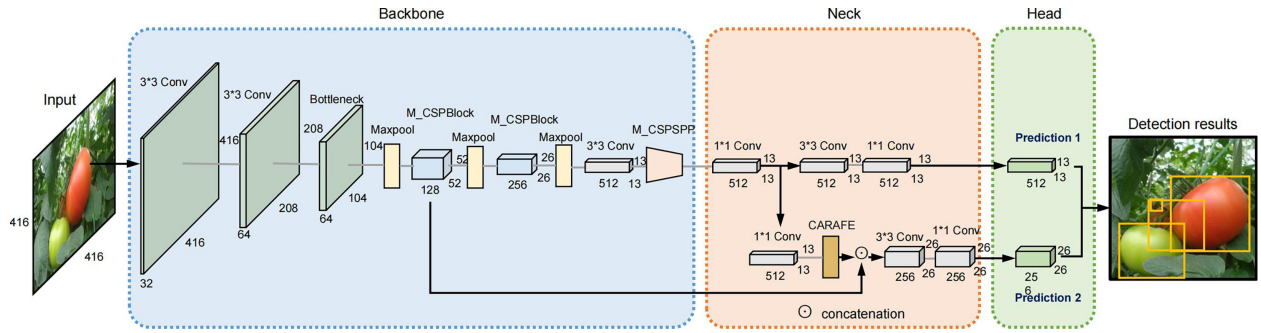


FIGURE 1
An overview of the improved YOLOv4-tiny model.

outperformed the original YOLOv4-tiny model and other state-of the-art object detectors in terms of accuracy (mAP (0.5)) and reached a real-time detection speed.

2 Theoretical background

2.1 YOLO series

YOLO is a state-of-the-art in real time object detection methods. YOLOv1, YOLOv2, and YOLOv3 (Redmon et al., 2016; Redmon and Farhadi, 2017; Redmon and Farhadi, 2018) were the first versions, and YOLOv2 was proposed with the objective of increasing the accuracy significantly. The idea of anchors for detection introduced in YOLOv2 was inspired by Faster R-CNN. It was also based on some other concepts such as Batch Normalization (Ioffe and Szegedy, 2015) and Skip connection (He et al., 2016). YOLOv3 evolved from YOLOv1 and YOLOv2 and became one of the state-of-the-art methods for object detection. It used Darknet-53 (Redmon and Farhadi, 2018) as a backbone instead of Darknet-19 (Redmon and Farhadi, 2017), multi-scale feature extractors (FPN) (Tsung-Yi et al., 2017), and binary cross-entropy loss instead of Softmax classification loss. YOLOv4 (Boschkovskiy et al., 2020) was released with the aim of improving YOLOv3.

Unlike Faster R-CNN, YOLO uses a different approach by applying a single neural network to a full image. This network divides the input into an $S \times S$ grid and performs detection in each cell. Each cell predicts bounding boxes along with the confidence of those boxes. These confidence scores reflect how confident the model is about whether the box contains an object or not. If it is confident, the confidence score tells how accurate the IoU of the ground truth (GT) and the predictions ($pred$) is. Equation (1) gives the formula of confidence:

$$Confidence = P(Object) \times IoU(GT, pred) \quad (1)$$

Where $P(Object) \in [0,1]$.

In YOLO model detection, each grid cell predicts C class probabilities for the object, so (5+ C) values are predicted by each cell: x, y, h, w, $Confidence$, and C class probabilities. x and y are the center coordinates of the box, and w and h are the width and the height of the box, respectively.

2.2 YOLOv4-tiny architecture

YOLOv4-tiny is a lightweight version of YOLOv4 that makes the network structure simpler and reduces parameters. It can achieve real-time detection. It uses a CSPDarknet-19 (Boschkovskiy et al., 2020) network as a backbone network instead of CSPDarknet-53, which is used in YOLOv4. By removing the computational bottlenecks that have a higher amount of calculation in the CSP-block module, it reduces the amount of calculation while increasing the accuracy. YOLOv4-tiny uses the LeakyReLU function as an activation function to simplify the computation process. Batch Normalization (BN) and Maxpooling are used between the layers of the CNN to speed-up training and select the maximum pixel values of features, respectively.

In the neck, a Feature Pyramid Network (FPN) is used. It can integrate different scales for implementing rich semantic information of a deep network and geometric detail of a shallow network to strengthen the ability of features extractions and to increase the object detection speed. The YOLO head uses features obtained by the FPN to make the final prediction and to form two prediction scales of 13×13 and 26×26.

2.3 Content-aware reassembly of features (CARAFE)

CARAFE (Jiaqi et al., 2019) is a feature map up-sampling operator that has two modules: a kernel prediction module and content-aware module. The kernel prediction module is responsible for generating the reassembly kernel in a content-aware manner. Each source location in the input corresponds to the target location σ^2 in the output. Each target location requires a $k_{up} \times k_{up}$ reassembly kernel, where k_{up} is the reassembly kernel size. It will output the reassembly kernels of size $C_{up} \times H \times W$, where $C_{up} = \sigma^2 \times k_{up}^2$.

The kernel prediction module has three sub-modules:

- A channel compressor sub-module reduces the channel of the input feature map by using a convolution layer (from C to C_m) with kernel size of 1×1.

- A content encoder sub-module takes the compressed feature map as input and encodes it to generate reassembly kernels by using a convolution layer of size $k_{encoder}$. The parameters of the encoder are $k_{encoder} \times k_{encoder} \times C_m \times C_{up}$.

- A kernel normalizer sub-module uses a Softmax function on each reassembly kernel.

The content-aware reassembly module reassembles the features within a local region *via* the function $\varnothing$. This function is just a weighted sum operator. For a target location l′ and a corresponding square region $N(Input_l, k_{up})$ centered at $l = (i, j)$, the reassembly is shown in Equation (2):

$$Output_{l'} = \sum_{n=-r}^{r} \sum_{m=-r}^{r} W_{l'_{(n,m)}} \cdot input_{(i+n,j+m)} \quad (2)$$

where $= \frac{k_{up}}{2}$, $W_{1'}$ is the location-wise kernel for each location l‘ based on the input, and l‘ is the neighbor location of l. The semantics of the reassembly feature maps is stronger with CARAFE than the original up-sampling operator because the information from relevant points in a local region is attended. CARAFE has several advantages: a large field of view, a content-aware handling, and it is lightweight and fast to compute.

3 Materials and methods

3.1 Image acquisition

The tomato datasets (Liu et al., 2020) used in this research were taken from December 2017 to November 2019 in Vegetable High-Tech Demonstration Park, Shouguang, China (36°51’44.2”N and 118°49’27.3”E). The images were taken using a digital camera (Sony DSC-W170, Tokyo, Japan) with a resolution of 3648 × 2056 pixels. The camera has a precision 5× wide-angle zoom Carl Zeiss Vario-Tessar lens with a range equivalent to a 28-140mm zoom on a 35mm camera, which allows it to take shots in tight spaces or get an entire group of things in the frame. Moreover, it incorporates Sony’s Super Steady-shot optical image stabilization to minimize blur caused by camera shake at slow shutter speeds. All the images were taken in natural daylight with different conditions including illumination variation, occlusion, and overlap. A total of 966 images were taken and divided into a training set and a test set. The training set had 725 images and contained 2553 tomatoes, while the test set had 241 images and contained 912 tomatoes. Figure 2 shows some examples from the datasets under different conditions.

3.2 Image augmentation

To prevent non-convergence phenomenon or over fitting during the training process, in this study, the images were augmented using various data-augmentation methods, such as rotation, noise, brightness transformation, and cutout, as shown in Figure 3 (Huang et al., 2020; Wu et al., 2020). To help the model to be insensitive to camera orientation, the original images were rotated by 90° and 270°. For the noise, we generated “salt and pepper” noise on the images, which can help the model to be more robust to noise. For the brightness transformation, we randomly changed the intensity of the pixels from -70% to 70%. Finally, a cutout method was adopted to help the model to be more resilient to object occlusion. All these methods were used before training to expand the datasets, which can help the model to be more accurate.

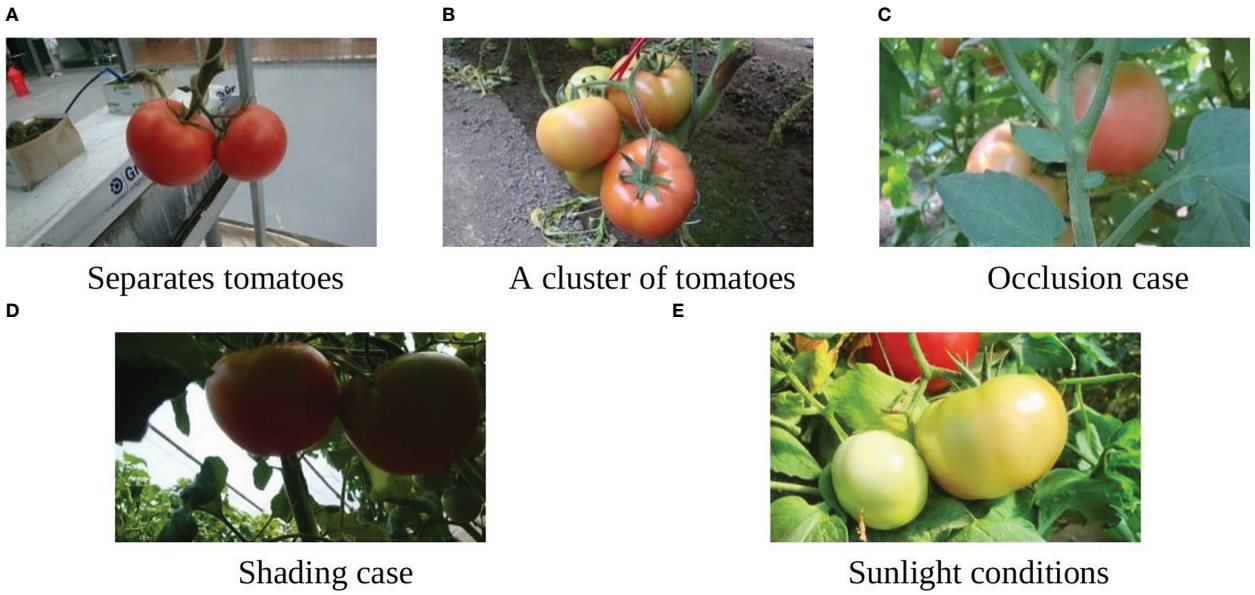


FIGURE 2
Tomato samples with different growing circumstances. **(A)** Separated tomatoes, **(B)** Cluster of tomatoes, **(C)** Occlusion case, **(D)** Shading case, and **(E)** Sunlight conditions.

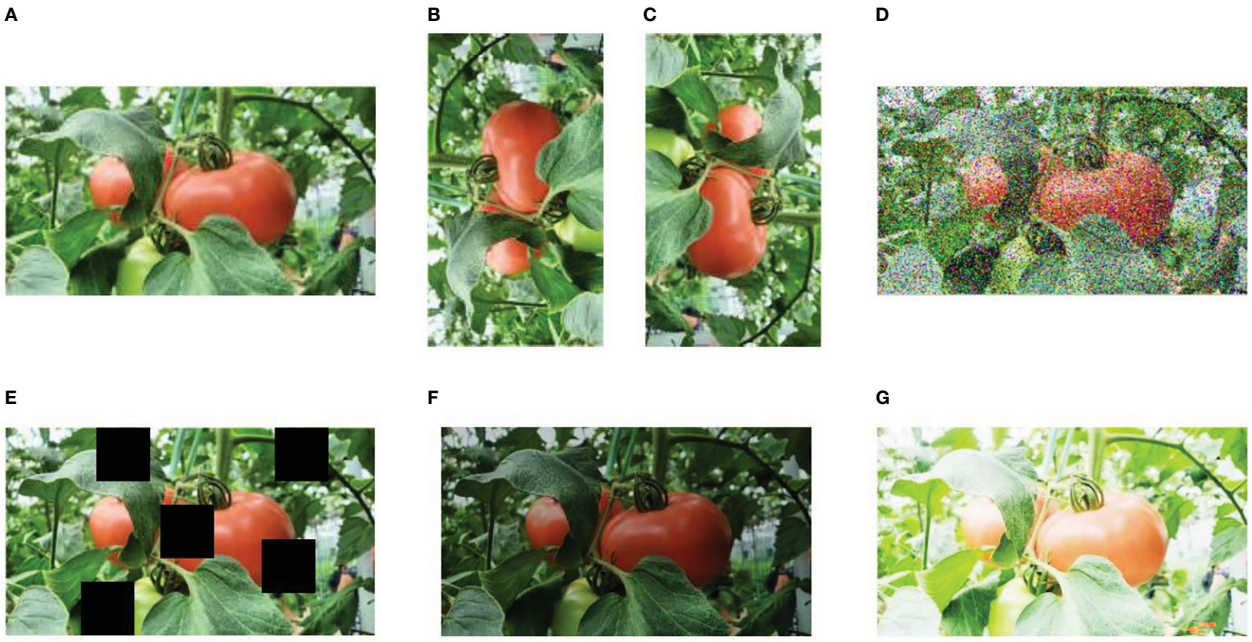


FIGURE 3
Some examples of image augmentation operations. **(A)** original image, **(B, C)** rotation (90° and 270°), **(D)** noise (salt and pepper), **(E)** cutout with five counts, and **(F, G)** exposure (brightness changes).

3.3 The improved YOLOv4-tiny model architecture

One of the advantages of YOLOv4-tiny is the fast detection speed because of its simplicity. However, due to the reduction of the number of layers, the feature capability is insufficient, and the feature utilization of the algorithm is low. This leads to low detection accuracy. To solve this issue, we propose a new model based on YOLOv4-tiny. The architecture of the improved model is shown in Figure 4.

3.3.1 The modified backbone network

To further improve the detection accuracy and robustness of YOLOv4-tiny model under complex conditions, it is needed to improve the detection accuracy further. The backbone of YOLOv4-tiny contains three BottleneckCSP modules, which consist of multiple convolutional layers, as shown in section 2.2. Even though the convolution operation can extract the features in the image, the convolutional kernel has a large number of parameters, which increases the computation load.

To reduce the number of parameters, the first BottleneckCSP of the original network is replaced with a Bottleneck module (He et al., 2016). Moreover, the original Bottleneck CSP module is modified to enhance feature extraction, capture more information, and reduce the computational complex. The modified BottleneckCSP is simpler, faster, lighter, and has better fusion characteristics.

The convolutional layer on the bridge branch of the original module was removed so that part of the input of the BottleneckCSP is directly connected to the output feature map of the other branch. This effectively reduces the number of parameters in the module. Figures 5, 6 show the bottleneck architecture and the difference between the original BottleneckCSP and the modified one, respectively.

From the original architecture of BottleneckCSP shown in Figure 6A, Equations (3) – (5) can be derived:

$$I_0 = \{I_{0'},\ I_{0''}\} \tag{3}$$

where I_0 is the input data, $I_{0'}$ is the first half of the input data, and $I_{0''}$ is the second half of the input data.

$$I_T = [Bottleneck(I_{0'}), Conv2d_{1\times1}(I_{0''})] \tag{4}$$

where I_T is the concatenate layer of $I_{0'}$ and Bottleneck of $I_{0''}$, and $Conv2d_{1\times1}$ is a convolutional layer with 1×1 kernel size.

$$Y = Conv2d_{1\times1}(LeakyReLu, BN(I_T)) \tag{5}$$

where Y is the output layer.

Similar to Equation (4), the concatenate layer of I_T in the new BottleneckCSP module is represented in Equation (6).

$$I_T = [Bottleneck(I_{0'}), I_{0''}] \tag{6}$$

The remaining two original BottleneckCSPs are replaced with the modified one in the backbone network to make it more efficient and enhance feature extraction.

Scales-YOLOv4 (Wang CY et al., 2021) introduced a CSP-Spatial Pyramid Pooling module (CSP-SPP), which used a cross stage process for down-sampling convolution operations. However,

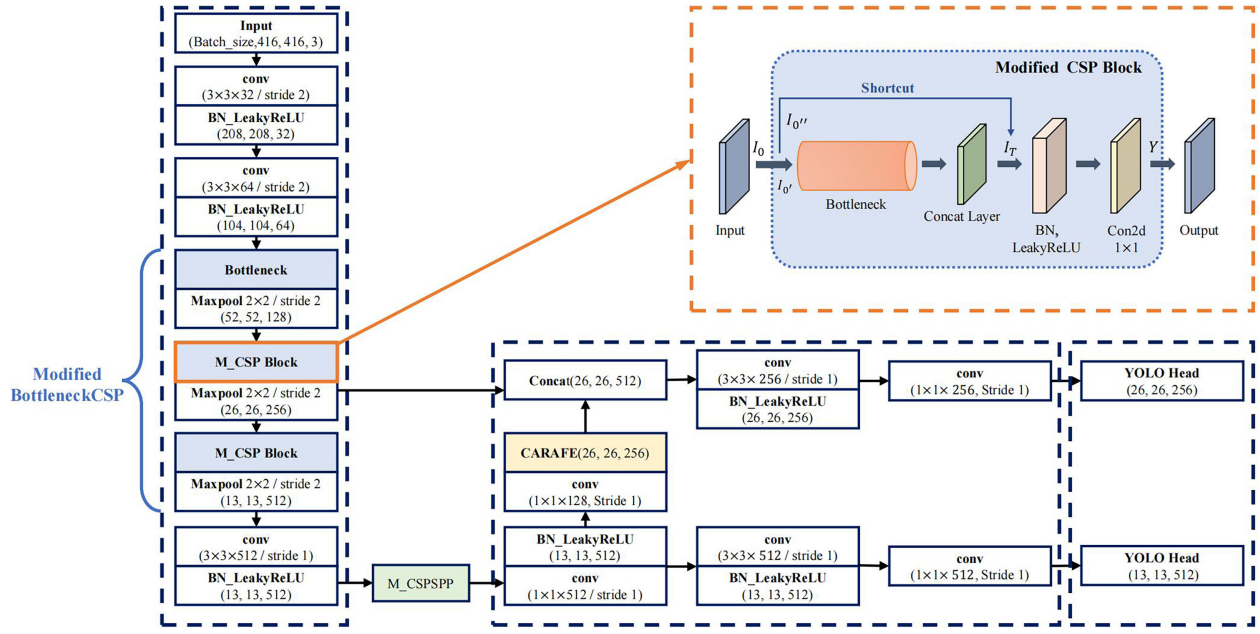


FIGURE 4
The improved YOLOv4-tiny model architecture.

it was designed for large-scale object detection models with large numbers of parameters and is not suitable for a tiny object detection. To adapt it to a tiny object detection, a tiny version of the CSP-Spatial Pyramid Pooling module is proposed in this study. It removes 1×1 and 3×3 convolutional layers to reduce the parameters and increase the accuracy of the model. Figure 7 shows the architecture of the original CSP-SPP and the tiny version of the module.

Equations (7) – (10) can be derived from the new module shown in Figure 7B:

$$f_1 = Conv_{1\times1}(I_0/_2) \tag{7}$$

$$f_2 = Maxpooling(f_1) \tag{8}$$

$$f_3 = Conv_{1\times1}([f_1,\ f_2]) \tag{9}$$

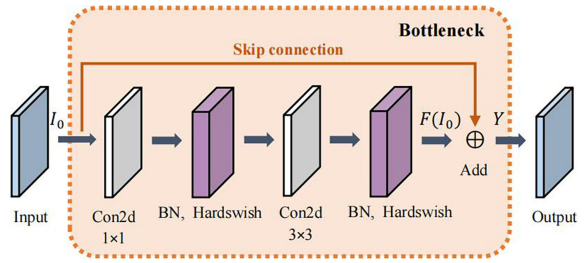


FIGURE 5
Bottleneck module architecture.

$$Y = Conv_{1\times1}([I_0,\ f_3] \tag{10}$$

Where $Conv_{1\times1}$ is a convolutional layer with 1×1 kernel size, f_1, f_2, and f_3 are feature maps, and Y is the output layer.

3.3.2 The modified neck network

In the YOLOv4-tiny neck, FPN (Tsung-Yi et al., 2017) is used to construct a feature pyramid of strong semantics with a top-down pathway and lateral connections. In the top-down pathway, a low-resolution feature map is firstly up-sampled twice with the nearest neighbour interpolation and then fused with a high-resolution one. It adopts spatial distance between pixels to guide the up-sampling process, but it considers only sub-pixel neighbours and fails to capture the rich semantic information required by dense prediction tasks. In Pixel shuffle (Shi et al., 2016) up sampling method, the feature map is extracted using sub-pixel convolution and then expands by a dimensional space. However, it scales the image size without changing the current amount of feature information. To solve this issue, all feature levels is substituted with CARAFE (Jiaqi et al., 2019), as shown in section 2.3. Figure 8 shows the new architecture. CARAFE is a region content-based up sampling method that first gets the up-sampling kernel in the up-sampling kernel prediction module, and uses it to up sample the corresponding positions of the original map. Then the new feature is used in the feature reassembly module to complete the up-sampling process and gets better output feature with high resolution. In addition, the kernel prediction module normalizes the features in the up-sampled region to maintain a constant value after up sampling, thereby

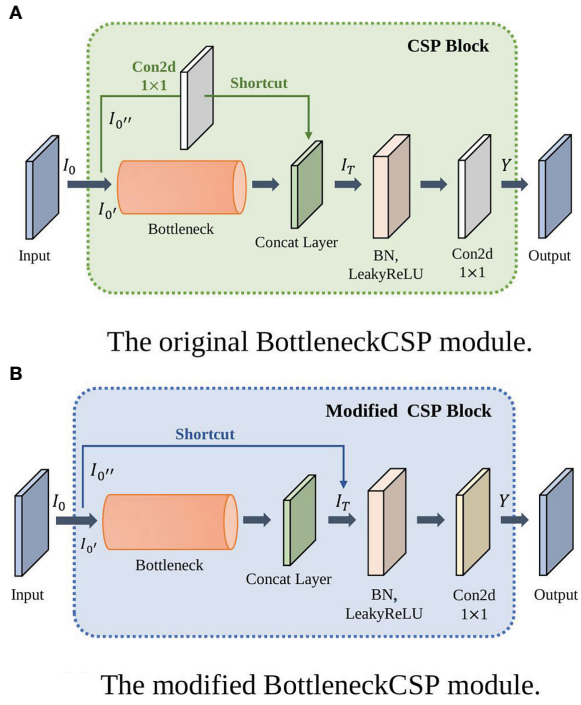


FIGURE 6
The BottleneckCSP module architectures. **(A)** Original BottleneckCSP, **(B)** The modified module.

reducing distortion. This modification is smooth, and no extra change is required. Moreover, it occupies less computing power, is lighter, and has demonstrated good performance in object detection and semantic segmentation tasks.

3.4 Experimental setup

In this study, the computer used had Intel i5, 64-bit 3.30-GHz quad-core CPUs (Santa Clara, CA USA), 16 GB of RAM, and an NVIDIA GeForce GTX 1070Ti GPU. The model framework was Pytorch with related software CUDA 11.1 and Python 3.8.10. The batch size was set to 8. The input image size was: 416 × 416. The setting of some hyper parameters used in this study is given as follows: number of epochs: 400, learning rate: 0.001, optimizer weight decay: 94.75, STD momentum: 96.3, warm-up initial momentum: 0.8, batch size: 8, box loss gain: 0.05, classification loss gain: 0.5, cls BCE loss positive weight: 1.0, object loss gain: 1.0, and anchor multiple threshold: 4.0.

3.4.1 Evaluation metrics

Evaluation indicators (Padilla et al., 2020) such as precision, recall, mAP, and F_1 score were used to evaluate the model performance. The indicators are defined as follows:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

where TP, FN, and FP are abbreviations for true positive (correct detection), false negative (miss), and false positive (false detection), respectively. The mAP was adopted to show the overall performance of a model under different confidence thresholds. It is defined as follows:

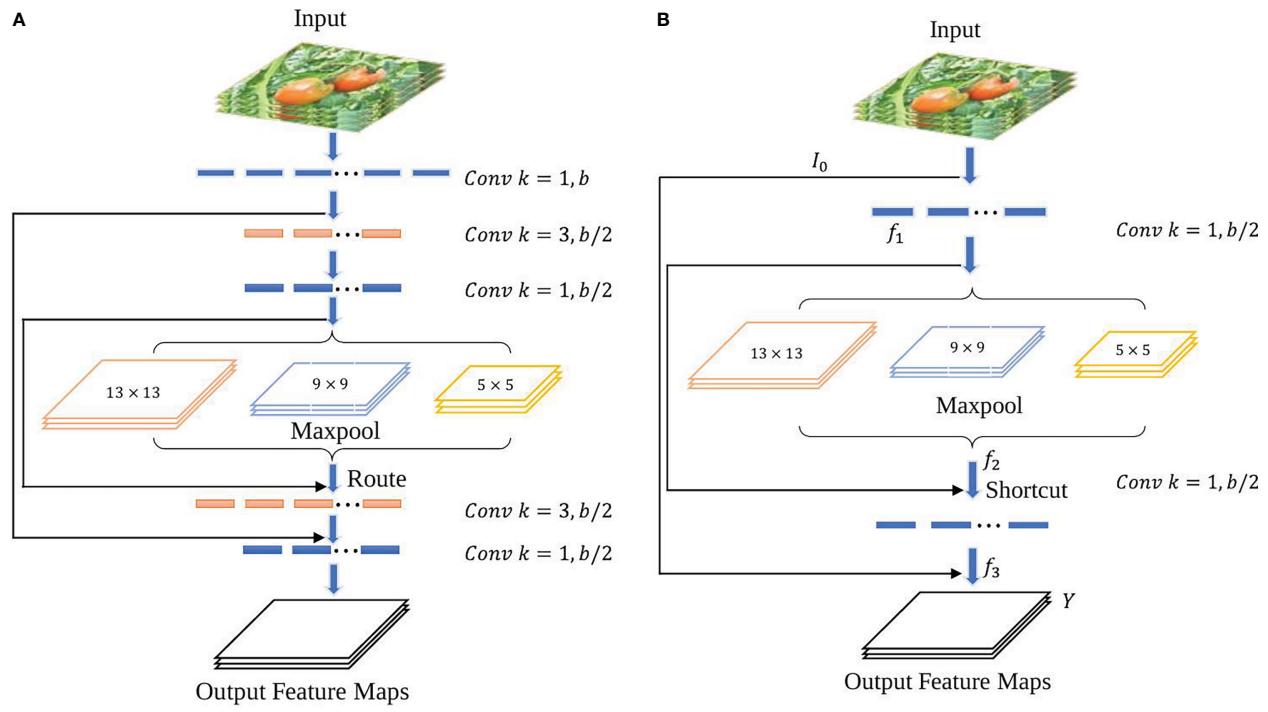


FIGURE 7
CSP-SPP module architectures. **(A)** the original module used in Scale-YOLOv4, **(B)** The tiny version of the CSP-SPP module.

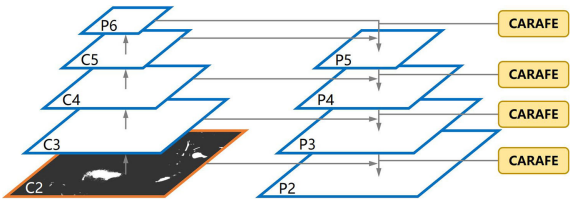


FIGURE 8
The FPN architecture with CARAFE.

$$mAP = \frac{1}{N_{classes}} \sum_{a=1}^{N_{cls}} AP_a \tag{13}$$

with

$$AP = \sum_{q}^{Q} (r_{q+1} - r_q) m\,axp(\tilde{r}) \tag{13.a}$$

where $p(\tilde{r})$ is the measured precision at recall $\tilde{r}$, and N_{cls} is the number of classes. The F_1 score is defined as follows:

$$F_1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{14}$$

3.4.2 Loss function

The loss function in this study considered the regression error of bounding coordinates, the confidence error of bounding box, and the classification error of object category. Equation (15) below shows how we calculated the loss function:

$$Loss = Loss_{reg} + Loss_{conf} + Loss_{cls} \tag{15}$$

- Loss regression:

$$Loss_{reg} = 1 - IoU + \frac{d^2\ (\hat{b}, b_{gt})}{c^2} + \alpha v \tag{15.a}$$

with

$$IoU = \frac{\hat{b} \cap b_{gt}}{\hat{b} \cup b_{gt}} \tag{15.b}$$

and

$$v = \frac{4}{\pi^2} (tan^{-1} \frac{w_{gt}}{h_{gt}} - tan^{-1} \frac{w}{h})^2,$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{15.c}$$

where b and b_{gt} are predicted bounding boxes and ground truth bounding boxes, respectively, d is the distance between the predicted center point and the true center point, c is the diagonal length of the enclosing box covering b and b_{gt}, and α and v are the positive trade-off and aspect ratio parameter, respectively.

From the equations above, we can see that the loss regression function works from three aspects: the overlap area, centroid distance, and the aspect ratio between the bounding box and the ground truth.

- Loss confidence:

To know the confidence loss, we need to calculate the confidence of the grid cell.

$$C = P(object) \times IoU(b, b_{gt}) \tag{15.d}$$

then,

$$Loss_{conf} = \sum_{i=1}^{s \times sNB} \sum_{j=1} \lambda_{i,j} [C_i \dot{n} log(\tilde{C}_i) log(1 - C_i)]$$

$$- \sum_{i=i}^{s \times sNB} \sum_{j=1} (1 - \lambda_{i,j}) [C_i \cdot \log \tilde{C}_i + (1 - C_i) \log (1 - \tilde{C}_i] \tag{15.e}$$

with

$$\lambda_{i,j} = \begin{cases} 1, & if\ part\ of\ j-th\ bounding\ box\ is\ in\ the\ i-th\ grid\ cell \\ 0, & otherwise \end{cases} \tag{15.f}$$

where s × s is the grid cell size, NB is the number of bounding boxes, $\tilde{C}_i$ is the obtained confidence from prediction box, and C_i is the confidence threshold.

- Loss classification:

$$Loss_{cls} = \sum_{i=1}^{s \times sNB} \sum_{j=1} \lambda_{i,j} \sum_{a \in classes} [p_i(a) \log (\tilde{p}_i(a)) + (1 - p_i(a)) \log (1 - \tilde{p}(a))] \tag{15.g}$$

where p_i is the true probability of detecting the object, $\tilde{p}_i$ is the probability score from the prediction, and a is a class associated with target detection. The loss function of YOLOv4-tiny converged gradually in the training process, such that the position and confidence of the bounding box are close to the ground truth.

4 Results and discussions

4.1 Ablation study

In this study, three major modifications were studied before obtaining the final result. Table 1 shows the ablation analysis of the different modifications. An ablation analysis of the impact of different modifications to the original YOLOv4-tiny was performed. Table 1 shows exactly what modifications were made.

First, the modified BottleneckCSP was incorporated into the backbone instead of the original BottleneckCSP module, which increased the accuracy by 1.7% and reduced the time by 0.9 ms compared to the original YOLOv4-tiny. Second, the tiny CSP-SPP module was attached to the modified backbone, which contributed another 1.9% improvement to the accuracy, and the time was reduced by 0.7 ms. Lastly, when the CARAFE module was adopted in the neck, the accuracy was further increased by 0.8%.

Also, we tested the function of the CARAFE module based on the modified BottleneckCSP and found that it improved the accuracy by 1.5%, which is a little lower than that of the CSP-SPP module. This showed the efficiency and effectiveness of each

TABLE 1 Ablation analysis of the different modifications.

	Modified BottleneckCSP	Modified CSP-SPP	CARAFE	mAP (0.5:0.95) (%)	Time (ms)
Modification				78.4	3.5
	✓			80.1	2.6
	✓	✓		82.0	1.9
	✓		✓	81.6	2.3
	✓	✓	✓	82.8	1.9

modification. In accordance with Table 1, Figure 9 shows that the accuracy and speed were both improved with the proposed modification.

Moreover, experiment was also performed with Pixel shuffle up sampling method (Shi et al., 2016). It implements sub-pixel method convolution to extract the feature map and then expands it by a dimensional space to obtain the up-sampling results. Compared with CARAFE, the Pixel shuffle has an accuracy of 81.03% and a detection time of 2.2ms, which are both worse than that of CARAFE.

4.2 Feature map visualization

Features were visualized in some stages of the algorithms (original YOLOv4-tiny and the improved YOLOv4-tiny). Figure 10 focuses on features where the original model was modified. Figure 10A shows an input image with tomatoes labeled for better visualization. Figures 10B, C represent stage 2 of both the original algorithm (the first BottleneckCSP module) and the modified algorithm (Bottleneck module), respectively. Figure 10D shows the second feature of the modified CSP-SPP module.

Stage 4 is shown in Figures 10E, F and represents the second BottleneckCSP module of the original algorithm and the modified Bottleneck module. Finally, Figures 10G, H show features in the original up-sampling operator in the neck and the CARAFE operator module, respectively. Moreover, CARAFE has a large field of view and can effectively aggregate context information, resulting in a good feature map. It can be seen in Figure 10H. Combining all the visualization in Figure 10, each modification has better features with high resolution than features in the original algorithm, which means that the improved model is better and more efficient than the original model.

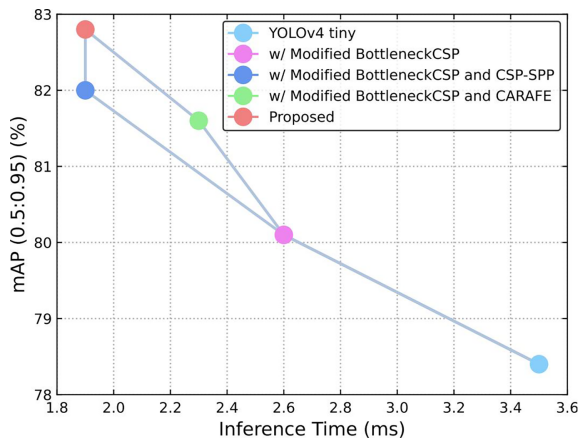


FIGURE 9
Speed (ms) versus accuracy (mAP).

4.3 Comparison of the improved YOLOv4-tiny with different one-stage detection algorithms

The performance of the improved YOLOv4-tiny was compared with other one stage detection algorithms: MobileNetv1 (Andrew et al., 2017), YOLOv3-tiny (Redmon and Farhadi, 2018), ShuffleNetv2 (Ma et al., 2018), MobileNetv3 (Howard et al., 2019), and YOlOv4-tiny (Boschkovskiy et al., 2020). Table 2 shows that the improved YOLOv4-tiny model has the best detection performance among all the methods. The mAP (0.5:0.95) was 7.4%, 11.5%, 6.2%, 5.4%, 0.8% and 4.4%, higher than those of Mobilenetv1, YOLOv3-tiny model, ShuffleNetv2, MobileNetv3, YOLOv5s, and YOLOv4-tiny model, respectively. The average detection time of the improved method was 1.9 ms, which met the requirement of real-time fruit detection.

As shown in Table 2, compared with MobileNetv1, YOLOv3-tiny, ShuffleNetv2, and YOLOv4-tiny, the precision of the improved model increased by 1.2%, 1.2%, 2.2% and 1.0%, respectively. However, the recall increased by 3.9%, 3.1%, 2.2%, and 1.0%, respectively. MobileNetv3 had almost the same precision with the improved model, whereas recall decreased by 4.1%. The F1 score and the mean average precision with IoU of 0.5 increased by 1.0% and 0.5% compared with that of the original YOLOv4-tiny model. Although the detection time of the improved model was slightly lower than that of MobileNetv1 and shuffleNetv2, the improvement of his accuracy was far better than that of MobilenetV1 and ShuffleNetv2. Moreover, the mAP (0.5:0.95) of the improved model is 0.8% higher than the one of YOLOv5s model, with less detection time. Performance of the Improved Model under Different Conditions

To evaluate the performance of the improved YOLOv4-tiny model under different lighting and occlusion environmental conditions, the tomatoes were divided into different groups.

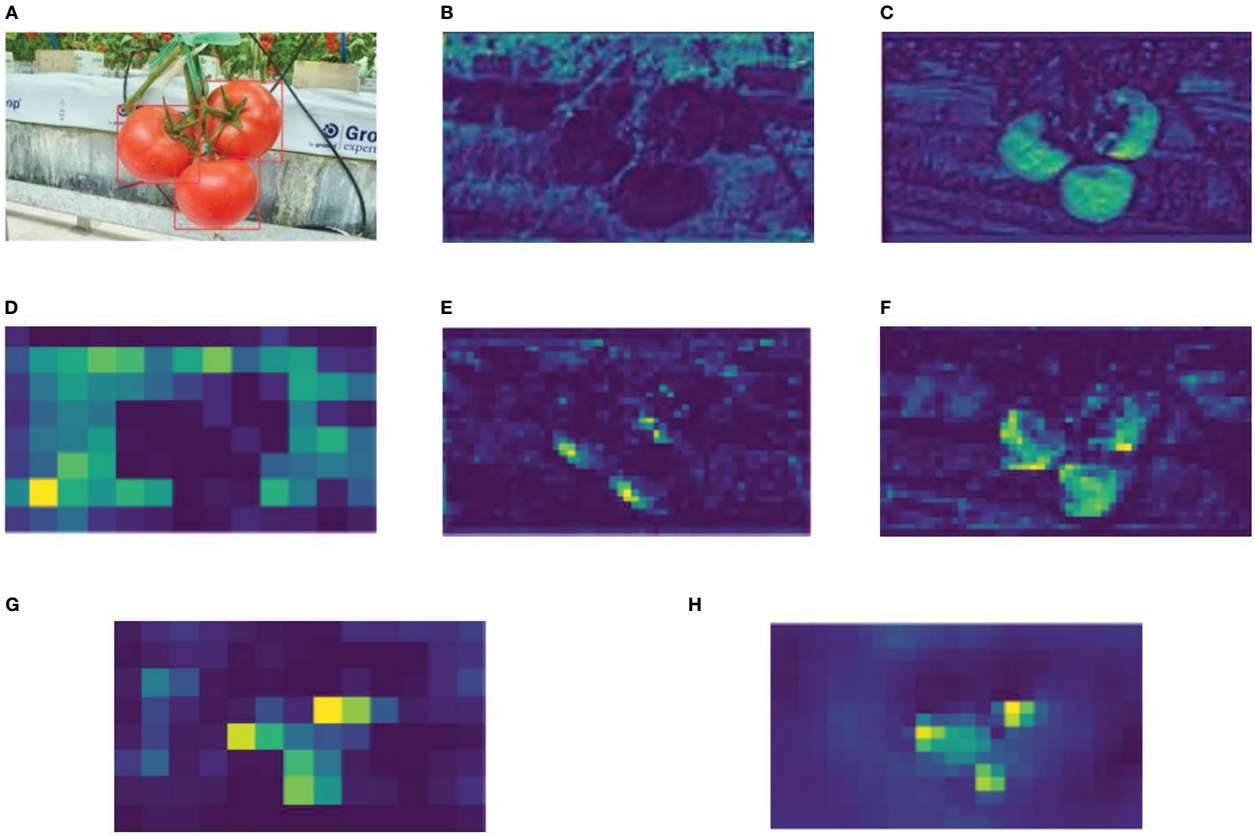


FIGURE 10
(A) The labeled input image, **(B)** 29th feature of the first BottleneckCSP of YOLOv4-tiny, **(C)** 29th feature of the Bottleneck module of the improved method, **(D)** 2nd feature of the modified CSP-SPP module, **(E)** 2nd feature of the second BottleneckCSP, **(F)** 2nd feature of the modified BottleneckCSP, **(G)** 29th feature of original up-sampling operator, **(H)** 29th feature of CARAFE operator.

According to different lighting conditions, the tomatoes were divided into sunlight and shading groups. Among all the 912 tomatoes, 487 of them are in sunlight conditions and 425 of them are in shading conditions. According to the degree of occlusion or overlap conditions, the tomatoes were divided into slight and severe occlusion cases. Severe cases refer to tomatoes being occluded by leaves, branches or other tomatoes by more than 50% degree.

Table 3 shows the evaluation results of the improved model under sunlight and shading conditions. 95.1% of the tomatoes were correctly detected under sunlight conditions while 94.8% for shading cases. The missed rates are 4.9% and 5.2% for sunlight and shading cases, respectively. Moreover, the false identification rates are 3.7% and 3.6% for sunlight and shading cases. This means that some leaves, branches or other background are falsely detected as tomatoes, especially when some background presents both similar color and shape as tomatoes.

Similarly, Table 4 shows the evaluation results of the improved model under slight and severe occlusion cases. Under slight occlusion case, 95.2% of the tomatoes were correctly detected, and 94.4% were detected under severe occlusion case. The results

TABLE 2 A comparison of the different models.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (0.5) (%)	mAP (0.5:0.95) (%)	GFLOPs	Time (ms)
MobileNetv1*	95.1	91.1	93.0	96.5	75.4	3	1.7
YOLOv3-tiny	95.1	91.9	93.4	97.4	71.3	10	3.8
ShuffleNetv2*	94.1	92.8	93.4	96.6	76.6	7	1.7
MobileNetv3*	96.4	90.9	93.5	96.8	77.4	8	1.6
YOLOv5s	96.3	94.2	95.2	98.3	81.7	16.8	2.7
YOLOv4-tiny	95.3	94.0	94.6	98.0	78.4	6.8	3.5
The Improved Yolov4-tiny	96.3	95.0	95.6	98.5	82.8	9	1.9

*MobileNetv1, Shufflenetv2, and MobileNetv3 were used as backbone network and YOLOv4-tiny head was used for detection.

TABLE 3 Performance of the improved model under different lighting conditions.

Conditions	Tomato Count	Correctly Identified		Falsely Identified		Missed	
		Amount	Rate (%)	Amount	Rate (%)	Amount	Rate (%)
Sunlight	487	463	95.1	18	3.7	24	4.9
Shading	425	403	94.8	15	3.6	22	5.2

TABLE 4 Performance of the improved model under different occlusion conditions.

Conditions	Tomato Count	Correctly Identified		Falsely Identified		Missed	
		Amount	Rate (%)	Amount	Rate (%)	Amount	Rate (%)
Slight occlusion	609	580	95.2	19	3.2	29	4.8
Severe occlusion	303	286	94.4	14	4.7	17	5.6

show that most of the tomatoes could be detected by our improved model except that are severely occluded by other objects. For the false identification rate, the results are 3.2% and 4.7%, respectively.

4.4 Qualitative analysis of different one-stage models

Figure 11 shows the prediction images of the comparison models, respectively. As shown in Figure 11, compared to the improved YOLOv4-tiny model, the other detection models have some either missed detections or false detections.

Moreover, the detection performance of the improved YOLOv4-tiny model was better and more efficient than that of the other detection models. The mean average with IoU of 0.5 to 0.95 increased by 4.4% compared to that of the original YOLOv4-tiny model, and the detection time per image was reduced by 1.6 ms. This means that the improved model is more accurate, compact and efficient for fruit detection in complex environment.

4.5 Comparison of the improved model with two-stage detection models

The performance of the improved model was compared with that of Faster R-CNN (Ren et al., 2015) and Dynamic R-CNN (Zhang et al., 2020), which are two-stage detection algorithms. Table 5 shows that Faster R-CNN took much time which led to huge amount of computation, whereas mAP with IoU of 0.5 to 0.95 of Faster R-CNN was 1.1% higher than that of the improved model. The detection time of the improved model was two time less than the detection time of Faster R-CNN. Moreover, the mAP with IoU of 0.5 to 0.95 of the improved model is 4.8% higher than that of the Dynamic R-CNN, with less time of detection. In summary, for the requirement of fruit detection which are accurate detection and a low amount of computation, the improved model is much better than the two-stage detection algorithm.

5 Conclusions and future work

To realize the application of tomato detection under complex environments, it needs a robust and efficient detection algorithm which is both accurate and fast. However, the existing methods are either inaccurate or slow for tomato detection, which cannot satisfy the requirement of tomato detection in the real natural environment. Thus, this study aims at proposing an efficient tomato detection algorithm based on YOLOv4-tiny, to obtain a more robust, fast and accurate tomato detection performance under complex environment conditions. To make the model more efficient, a modified backbone was proposed. The BottleneckCSP modules were replaced in the original backbone with a Bottleneck and modified BottleneckCSP modules to enhance feature extraction and reduce the computational complex. Moreover, a light version of the CSP–SPP module was attached to the modified backbone to improve the receptive field. Finally, to obtain a better feature map with high resolution, the traditional up-sampling operator in the neck was replaced by CARAFE.

Extensive experiments were conducted to verify the performance of the improved model. An ablation study proved the effectiveness of each modification. With the above modifications, the mAP (0.5:0.95) were increased by 1.7%, 1.9% and 0.8%, respectively, showing that the detection performance was greatly improved. The precision, recall, F1 score, mAP (0.5), and mAP (0.5:0.950) were 96.3%, 95.0%, 95.6%, 98.5%, and 82.8%, respectively. The detection speed reached 1.9 ms per image.

Furthermore, the performance of the improved method under different lighting and occlusion conditions were evaluated. The performance of the model was comparable under sunlight and shading conditions, showing that the model was robust to illumination variation. However, the model showed a divergence under different occlusion conditions. Under slight occlusion, 95.2% of the tomatoes were correctly detected, while 94.4% were detected under severe occlusion case. This showed that occluded and overlapped tomatoes could cause inaccurate detections, especially when the occlusion degree exceeds 50%.

10.3389/fpls.2023.1150958



FIGURE 11
Detection results of different models: **(A–E)** are the labeled images, **(F–J)** are prediction images from MobileNetv1, **(K–O)** are prediction image from YOLOv3-tiny, **(P–T)** are prediction images from ShuffleNetv2 detection, **(U–Y)** are prediction images from MobileNetv3 detection, **(A'–E')** are the prediction images from YOLOv5s detection, **(F'–J')** are the prediction images from YOLOv4-tiny detection, **(K'–O')** are the prediction images from the improved YOLOv4-tiny detection. (* MobileNetv1, ShuffleNetv2, and MobileNetv3 were used as backbone network and YOLOv4-tiny head was used for detection).

TABLE 5 A comparison of the improved model and two stage detection model.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (0.5) (%)	mAP (0.5:0.95) (%)	Time (ms)
Faster R-CNN(VGG-16)	96.5	94.8	95.6	97.8	83.9	3.9
Dynamic R-CNN	95.3	93.2	94.2	96.6	78.0	2.4
The Improved YOLOv4-tiny	96.3	95	95.6	98.5	82.8	1.9

The improved YOLOv4-tiny model was compared with some other state-of-the-art algorithms. The results showed that the improved model performed better than the other one-stage models. Moreover, the improved algorithm was compared with two-stage object detection algorithms (Faster R-CNN and Dynamic R-CNN). The results showed that the detection accuracy of the improved model could match that of the two-stage detection models and was faster. This indicates great potential of the improved model for tomato detection in complex environment.

In future work, based on the proposed model in this study, the information about tomato ripeness will be incorporated to classify a tomato in different growing stages. Moreover, further research will be conducted to improve the accuracy for severely occluded tomatoes.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

PLTM conceived the idea. PLTM and GL designed the methodology. PLTM and JS performed the experiments and analysis. PLTM wrote the original draft. JK and SK revised the manuscript. JK and GL supervised the experiments. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Weifang Science and Technology Development Plan (2021GX054), Doctoral Research Foundation of Weifang University (2022BS70).

Acknowledgments

This research was supported by BK21PLUS, Creative Human Resource Development Program for IT Convergence.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Afonso, M., Fonteijn, H., Fiorentin, F. S., Lensink, D., and Faber, N. (2020). Tomato fruit detection and counting in greenhouses using deep learning. *Front. Plant Sci.* 11, 1759. doi: 10.3389/fpls.2020.571299

Andrew, G. H., Menglong, Z., Bo, C., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*, arXiv:1704.04861. doi: 10.48550/arXiv.1704.04861

Arunabha, M. R., and Jayabrata, B. (2021). A deep learning enabled multi-class plant disease detection model based on computer vision. *AI* 2, 413–428. doi: 10.3390/ai2030026

Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint* arXiv:2004.10934 doi: 10.48550/arXiv.2004.10934

Chenglin, W., Wang, Y., Liu, S., Lin, G., He, P., Zhang, Z., et al. (2022). Study on pear flowers detection performance of YOLO-PEFL model trained with synthetic target images. *Front. Plant Sci* 12. doi: 10.3389/fpls.2022.911473

Cheng, Y., Zhang, B., Zhou, J., and Wang, K.. (2020). Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots. *Comput. Electron. Agric.* 175, 105579. doi: 10.1016/j.compag.2020.105579

Fu, L., Yali, F., Wu, J., Liu, Z., Gao, F., Majeed, Y., et al. (2020). Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precis. Agric.* 22, 754-776. doi: 10.1007/s11119-020-09754-y

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., and Karkee, M. (2020). Multi-class fruit-on-plant detection for apple in SNAP system using faster r-CNN. *Comput. Electron. Agric.* 176, 105634. doi: 10.1016/j.compag.2020.105634

Gao, H., Zhuang, L., Laurens, V. D. M., and Kilian, Q. W. (2018). Densely connected convolutional networks. In *Proceeding of the IEEE conference on computer vision and pattern recognition.*, 4700-4708. doi: 10.48550/arXiv.1608.06993

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceeding of the IEEE conference on Computer vision and pattern recognition.*, 580-587. doi: 10.48550/arXiv.1311.2524

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer vision and Pattern Recognition* (Las Vegas, NV, USA, IEEE), 770–778. doi: 10.1109/CVPR.2016.90

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., et al. (2019). Searching for MobileNetV3. In *Proceeding of the IEEE/CVF international conference on Computer vision*, 1314-1324. doi: 10.48550/arXiv.1905.02244

Huang, L., Pan, W., Zhang, Y., Qian, L., Gao, N., and Wu, Y. (2020). Data augmentation for deep learning-based radio modulation classification. *IEEE Access* 8, 1498–1506. doi: 10.1109/Access.2019.2960775

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, 448-456. doi: 10.48550/arXiv.1502.03167

Jiaqi, W., Kai, C., Rui, X., Ziwei, L., Chen, C. L., and Dahua, L. (2019). CARAFE: Content-Aware-Reassembly of features. *In proceedings of the IEEE/CVF international conference on computer vision*, 3007-3016. doi: 10.48550/arXiv.1905.02188

Kaiming, H., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask r-CNN. *In proceedings of the IEEE international conference on computer vision*, 2961-2969. doi: 10.48550/arXiv.1703.06870

Kamilaris, A., and Prenafeta-Boldu, F. X. (2018). Deep learning in agriculture: A survey. *Compters Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016

Kelman, E. E., and Linker, R. (2014). Vision-based localization of mature apples in tree images using convexity. *Biosyst. Eng.* 118, 174–185. doi: 10.1016/j.biosystemseng.2013.11.007

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks," in *Proceedings of the International Conference on Neural Information Processing Systems*, Lake Tahoe, NV, USA, Vol. 25. 1097–1105.

Linker, R., Cohen, O., and Naor, A. (2012). Determination of the number of green apples in RGB images recorded in orchards. *Comput. Electron. Agric.* 81, 45–57. doi: 10.1016/j.compag.2011.11.007

Liu, G., Mao, S., and Kim, J. H. (2019). A mature-tomato detection algorithm using machine learning and color analysis. *Sensors* 19, 2023. doi: 10.3390/s19092023

Liu, G., Nouaze, C. P., Touko, M. P. L., and Ho Kim, J. (2020). YOLO-tomato: a robust algorithm for tomato detection based on Yolov3. *Sensors* 20, 2145. doi: 10.3390/s20072145

Ma, N., Zhang, X., Tao-Zheng, H., and Sun, J. (2018). ShuffleNet V2: Practical guidelines for efficient CNN architecture design. *In Proceedings of the European conference on computer vision (ECCV)*, 116-131. doi: 10.48550/arXiv.1807.11164

Mu, Y., Chen, T. S., Ninomiya, S., and Guo, W. (2020). Intact detection of highly occluded immature tomatoes on plants using deep learning techniques. *Sensors* 20, 2984. doi: 10.3390/s20102984

Padilla, R., Netto, S. L., and Eduardo, A. B. (2020). "A survey on performance metrics for object detection algorithms," in *International Conference on Systems, Signals, Signals, and Image Processing (IWSSIP)*, Niteroi, Brazil, IEEE. 237–242. doi: 10.1109/IWSSIP48289.2020.9145130

Rahnemoofar, M., and Sheppard, C. (2017). Deep count: Fruit counting based on deep simulated learning. *Sensors* 17, 905. doi: 10.3390/s17040905

Redmon, J., and Farhadi, A. (2017). "YOLO9000: Better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Honolulu, HI, USA: IEEE) 7263-7271. doi: 10.1109/CVPR.2017.690

Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. doi: 10.48550/arXiv.1804.02767

Redmon, J., Farhadi, A., Divvala, S., and Girshick, R. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*. (Las Vegas, NV, USA: IEEE), 779–788. doi: 10.48550/arXiv.1506.02640

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Proceedings of the International Conference on Neural Information Processing Systems:* Montreal, Canada, Vol. 39. 91–99. doi: 10.1109/TPAMI.2016.2577031

Rupareliya, S., Gajjiar, R., and Jethva, M. (2022). "Real-time tomato detection, classification, and counting system using deep learning and embedded systems," in *Proceedings of the International e-Conference on Intelligent Systems and Signal Processing* Singapore, Springer, 511–522. doi: 10.1007/978-981-16-2123-9_39

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., and McCool, C. (2016). Deep fruits: A fruit detection system using deep neural networks. *Sensors* 16, 1222. doi: 10.3390/s16081222

Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, P. A., Bishop, R., et al. (2016). Real-time single image and video super-resolution using an efficient Sub-pixel convolutional neural network. *In Proceedings of the IEEE conference on Computer vision and pattern recognition.*, 874-1883. doi: 10.48550/arXiv.1609.05158

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. doi: 10.48550/arXiv.1409.1556

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the Thirty-first AAAI conference on Artificial Intelligence*, San Francisco, California, USA, AAAI Press. doi: 10.1609/aaai.v31i1.11231

Tang, Y., Zhou, H., Wang, H., and Zhang, Y.. (2023). Fruit detection and positioning technology for a camellia oleifera C. Able orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Systems with Applications* V211, 118573 doi: 10.1016/j.eswa.2023.118573

Tsung-Yi, L., Piotr, D., Ross, G., Kaiming, H., Bharath, H., and Serge, B. (2017). Feature pyramid networks for object detection. *In Proceedings of the IEEE conference on Computer vision and pattern recognition*, 2117-2125. doi: 10.48550/arXiv.1612.03144

Wang, C. Y., Boschkovskiy, A., and Liao, H. Y. M. (2021). Scaled-Yolov4: Scaling cross stage partial network. *In Proceedings of the IEEE/CVF conference on Computer vision and pattern recognition*, 13029-13038. doi: 10.48550/arXiv.2011.08036

Wang, C. Y., Boschkovskiy, A., and Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*. doi: 10.48550/arXiv.2207.02696

Wang, X., Liu, J., and Liu, G. (2021). Diseases detection of occlusion and overlapping tomato leaves based on deep learning. *Front. Plant Sci* 13. doi: 10.3389/fpls.2021.792244

Wu, X., Qi, Z., Wang, L., Yang, J., and Xian, X. (2020). Apple detection method based on light-YOLOv3 convolutional network. *Trans. CSAM* 51, 17–25.

Xu, Z., Jia, R., Liu, Y., Zhao, C., and Sun, H. (2020). Fast method of detecting tomatoes in a complex scene foe picking robots. *IEEE Access* 8, 55289-55299. doi: 10.1109/ACCESS.2020.2981823

Zhang, H., Chang, H., Ma, B., Wang, N., and Chen, X. (2020). "Dynamic r-CNN: Towards high quality object detection *via* dynamic training," in *European Conference on Computer Vision (ECCV)*. 260–275 (Cham: Springer). doi: 10.1007/978-3-030-58555-6_16

Zhao, Y., Gong, L., Huang, Y., and Liu, C. (2016). Robust tomato recognition for robotic harvesting using feature images fusion. *Sensors* 16, 173. doi: 10.3390/s16020173

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 April 2023
DOI 10.3389/fpls.2023.1108355

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
United States Department of Agriculture (USDA), United States

REVIEWED BY
Jun Yan,
China Agricultural University, China
Yong Suk Chung,
Jeju National University, Republic of Korea

*CORRESPONDENCE
Talukder Z. Jubery
znjubery@iastate.edu
Baskar Ganapathysubramanian
baskarg@iastate.edu

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 25 November 2022
ACCEPTED 28 March 2023
PUBLISHED 14 April 2023

CITATION
Dong D, Nagasubramanian K, Wang R, Frei UK, Jubery TZ, Lübberstedt T and Ganapathysubramanian B (2023) Self-supervised maize kernel classification and segmentation for embryo identification. *Front. Plant Sci.* 14:1108355. doi: 10.3389/fpls.2023.1108355

COPYRIGHT
© 2023 Dong, Nagasubramanian, Wang, Frei, Jubery, Lübberstedt and Ganapathysubramanian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Self-supervised maize kernel classification and segmentation for embryo identification

David Dong[1,2], Koushik Nagasubramanian[2,3], Ruidong Wang[4], Ursula K. Frei[4], Talukder Z. Jubery[2,5]*, Thomas Lübberstedt[4] and Baskar Ganapathysubramanian[2,3,5]*

[1]Ames High School, Ames, IA, United States, [2]Translational AI Center, Iowa State University, Ames, IA, United States, [3]Department of Electrical Engineering, Iowa State University, Ames, IA, United States, [4]Department of Agronomy, Iowa State University, Ames, IA, United States, [5]Department of Mechanical Engineering, Iowa State University, Ames, IA, United States

Introduction: Computer vision and deep learning (DL) techniques have succeeded in a wide range of diverse fields. Recently, these techniques have been successfully deployed in plant science applications to address food security, productivity, and environmental sustainability problems for a growing global population. However, training these DL models often necessitates the large-scale manual annotation of data which frequently becomes a tedious and time-and-resource- intensive process. Recent advances in self-supervised learning (SSL) methods have proven instrumental in overcoming these obstacles, using purely unlabeled datasets to pre-train DL models.

Methods: Here, we implement the popular self-supervised contrastive learning methods of NNCLR Nearest neighbor Contrastive Learning of visual Representations) and SimCLR (Simple framework for Contrastive Learning of visual Representations) for the classification of spatial orientation and segmentation of embryos of maize kernels. Maize kernels are imaged using a commercial high-throughput imaging system. This image data is often used in multiple downstream applications across both production and breeding applications, for instance, sorting for oil content based on segmenting and quantifying the scutellum's size and for classifying haploid and diploid kernels.

Results and discussion: We show that in both classification and segmentation problems, SSL techniques outperform their purely supervised transfer learning-based counterparts and are significantly more annotation efficient. Additionally, we show that a single SSL pre-trained model can be efficiently finetuned for both classification and segmentation, indicating good transferability across multiple downstream applications. Segmentation models with SSL-pretrained backbones produce DICE similarity coefficients of 0.81, higher than the 0.78 and 0.73 of

those with ImageNet-pretrained and randomly initialized backbones, respectively. We observe that finetuning classification and segmentation models on as little as 1% annotation produces competitive results. These results show SSL provides a meaningful step forward in data efficiency with agricultural deep learning and computer vision.

KEYWORDS

self-supervised, classification, embryo identification, segmentation, high-throughput sorting

1 Introduction

Deep learning (DL) for computer vision applications has recently become a boon to innovations in agricultural efficiency. These methods have transformed how we extract various agronomically relevant plant traits under laboratory and field conditions (Fahlgren et al., 2015; Ubbens and Stavness, 2017; Singh et al., 2018; Guo et al., 2021). Automatically and rapidly extracting plant traits can be a game-changer in terms of reducing food costs and improving production efficiencies, improving sustainability by reducing waste, and providing a better understanding of adapting crops for climate change. Deep learning methods have been used in various agricultural applications to identify, classify, quantify, and predict traits (Mohanty et al., 2016; Naik et al., 2017; Pound et al., 2017; Dobrescu et al., 2019; Jubery et al., 2021). With the availability of high-throughput data acquisition tools that produce large amounts of good-quality data, the major bottleneck in deploying DL-based computer vision tools is the need for large amounts of labeled data to train these DL models. Data annotation or labeling is the main development barrier to building high-quality DL models, especially since labeling the raw data often requires domain experts to annotate images. Data annotation by an expert with domain-specific knowledge is a tedious and expensive task. The DL community is exploring various strategies to break this dependency on a large quantity of annotated data to train DL models in a label-efficient manner, including approaches like active learning (Nagasubramanian et al., 2021), transfer learning (Jiang and Li, 2020), weakly supervised learning (Ghosal et al., 2019; Körschens et al., 2021) and the more recent advances in self-supervised learning (Jing and Tian, 2020; Marin Zapata et al., 2021; Nagasubramanian et al., 2022). Transfer learning has been widely utilized in plant phenomics applications for classification and segmentation tasks (Wang et al., 2019; Kattenborn et al., 2021). Recently, self-supervised learning has been applied to improve classification and segmentation models (Güldenring and Nalpantidis, 2021; Nagasubramanian et al., 2022; Lin et al., 2023). In this work, we focus on deploying self-supervised learning approaches to the problem of characterizing maize kernels that are imaged in a commercial high-throughput seed imaging system [Qsorter technologies (QualySense)]. We consider two vision tasks – first, identify if the maize kernels are correctly oriented for downstream analysis (a classification task), and second, segment out the kernel scutellum from the correctly oriented seeds (a segmentation task).

The ability to accurately and efficiently segment maize kernel scutellum has significant utility for both production and breeding application. Maize oil (corn oil) is extracted from corn kernels through milling (Paulsen and Hill, 1985). Milling processes are integrated into the production of corn starch, sugar, syrup, alcohol, and byproducts like gluten feed, along with corn oil. Of the 1.1 billion metric tons of corn produced annually around the world, over 3.5 million are used for oil production (Ward and Singh, 2002; Lee et al., 2021). Almost all oil is found in the embryo of the kernel (Paulsen and Hill, 1985). The ability to sort seeds for embryo/scutellum size is a significant value addition. Similarly, the non-destructive sorting of single seeds based on oil content (OC) has been shown to be useful for early-generation screening to improve the efficiency of breeding (Silvela et al., 1989; Xu et al., 2019) and for haploid selection in an oil-inducer-based doubled haploid breeding program (Chaikam et al., 2019; Aboobucker et al., 2022). Over the past few years, nuclear magnetic resonance (NMR) (Melchinger et al., 2017; Yang et al., 2018), fluorescence imaging (Boote et al., 2016), near-infrared (NIR) reflectance spectroscopy (Jiang et al., 2007; Armstrong et al., 2011; Jones et al., 2012; Gustin et al., 2020), hyperspectral imaging (Weinstock et al., 2006), and line-scan Raman hyperspectral imaging (Liu et al., 2022) have been developed to measure or predict oil content. However, these methods and tools are expensive. On the other hand, sorting based on NIR reflectance is less costly, has been around for a long time (McClure, 2003; Halcro et al., 2020), and has worked well to predict protein and starch content. However, using those tools to measure OC is not easy because the position of the embryo/scutellum to the camera/light source (Spielbauer et al., 2009) strongly affects OC measurements of single seeds, which leads to significant standard errors. Several currently available NIR spectra-based high throughput single seed sorting devices capture RGB images of the seed along with the NIR spectrum (QualySense; Satake-USA). These images can be used to identify the correctly orientated seed and quantify the relative size of the embryo to the seed, which, coupled with the NIR spectrum, could be used to improve the prediction of OC.

This work aims to design an end-to-end DL framework that classifies kernels based on their orientation and segments the embryos of correctly oriented kernels. Accurately performing these steps will allow us to, in the future, predict corn OC with high accuracy. Figure 1 illustrates this pipeline. A challenge in accomplishing this goal is that DL techniques often rely on having access to large datasets of annotated images for successful training results. This problem motivates our approach of using self-supervised contrastive. The self-supervised pretraining procedure automatically uses unlabeled data to generate pretrained labels (Misra and Maaten, 2020). It does so by solving a pretext task suited for learning representations, which in computer vision typically consists of learning invariance to image augmentations like rotation and color transforms, producing feature representations that ideally can be easily adapted for use in a downstream task. After obtaining this pre-trained model, we apply standard DL to finetune the model with a smaller labeled dataset. The smaller labeled dataset is used to reduce the effect of possible inaccuracies in the pseudo-labels from the self-supervised task (Zhai et al., 2019). The orientation of corn kernels must maintain consistency between measurements and be oriented to fully display the embryo. The goal of the segmentation problem is then to identify the embryo amidst the background and the rest of each kernel.

Our contributions in this paper are 1) the creation of an end-to-end DL pipeline for kernel classification and segmentation, facilitating downstream applications in OC prediction, 2) to assess capabilities of self-supervised learning regarding annotation efficiency, and 3) illustrating the ability of self-supervised pretraining to create models that can be finetuned for diverse downstream applications. Beyond the direct application of the classification and segmentation capabilities of the learned representations, using self-supervised techniques, in general, could accelerate the development of computer vision techniques for ag applications, skipping several stages of arduous and time-consuming data collection.

2 Materials and methods

2.1 Dataset

2.1.1 Dataset for classification by imaging orientation

The classification dataset consists of 44,286 RGB 492-pixel by 240-pixel images of maize kernels of various accessions taken using the RGB imaging tools of QSorter. Of these, 2697 were manually labeled into two classes: “oriented” and “non-oriented.” Kernels that belong to the “oriented” class were deemed appropriate for calculating internal OC within the embryo/germ center of corn kernels. This determination was based on the requirement that the visible embryo is parallel to the camera’s plane.

In a typical downstream application, this visual information provided by image segmentation would be combined with data from the hyperspectral imaging sensor provided by QSorter, but with such a sensor having its field of view limited to only the middle pane. However, the other two panes still provide useful visual information for our classification models since the determination of the orientation of any particular kernel is not limited to only the frontal view of the kernel. Figure 2A shows oriented kernels, noting

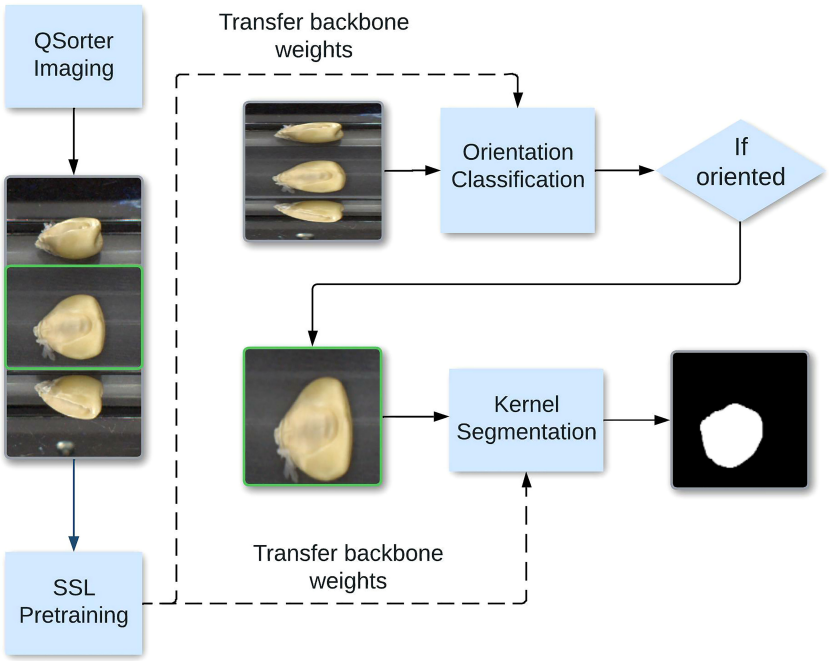


FIGURE 1
End-to-end pipeline for corn kernel classification and segmentation. The curved arrow shows the middle pane being processed for the segmentation task. We show that the classification and segmentation models perform bets with self-supervised weights.

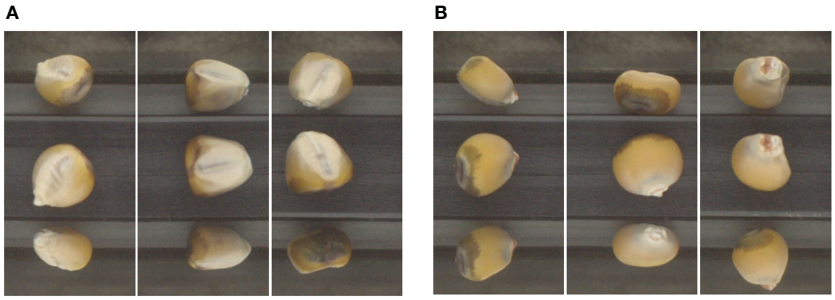


FIGURE 2
Images classified as "oriented" with the embryo visible **(A)** and "non-oriented" with the embryo not visible **(B)**.

the lighter portion visible in each middle pane, which is the corn embryo's visible part. Figure 2B shows non-oriented kernels in which the embryos are not visible or only partially visible.

2.1.2 Dataset for embryo segmentation

The embryo segmentation dataset consists of only 401 RGB images of corn kernels, taken from the same source of QSorter images as in the classification dataset above, along with their respective binary masks. Thus, the 2D image shapes were again 492 x 240. Segmentation (into the binary mask) distinguishes between the embryo and the rest of the background (including the non-embryo portion of the kernel). Figure 3 illustrates the segmentation annotation process for an RGB image and its mask. The three frames of each original (492, 240) dataset image were split into three individual images and downsampled to (128, 128). All completely negative masks and their respective RGB images were then removed.

2.2 SSL pretraining

2.2.1 Methods overview

The contrastive learning framework is a self-supervised learning method that maximizes the similarity between representations of an image and the augmented version of an image while minimizing the similarity between an image and other images (Zhao et al., 2021). The two models used for self-supervised pretraining were SimCLR (Simple Framework for Contrastive Learning of Visual Representations) (Chakraborty et al., 2020) and NNCLR (Nearest-Neighbor Contrastive Learning of Visual Representations) (Dwibedi et al., 2021). Figure S1 shows these two models superimposed on the same diagram.

SimCLR trains a backbone used for downstream processes by considering the contrastive loss of the representations of two distinct augmentations of images extracted from any given batch. If the initial images are the same, the pair of representations is considered a positive pair for the final calculation, and if the views are augmentations of two distinct images in the batch, then it is considered a negative pair. The representations are created by taking each augmented view of the initial image along a path including two networks: a base encoder where the desired backbone resides and a final projection head to calculate the contrastive loss of the representation in a separate space. NNCLR is also a contrastive model but differs from SimCLR in that upon taking both views of a given image through an encoder; the nearest neighbor algorithm is used to sample dataset representations for one of the views from a subset of the initial dataset. These are treated as the analog of the positive pairs described in the SimCLR model. Negative pairs are then the nearest neighbors of distinct initial images. Both architectures use the same InfoNCE loss to

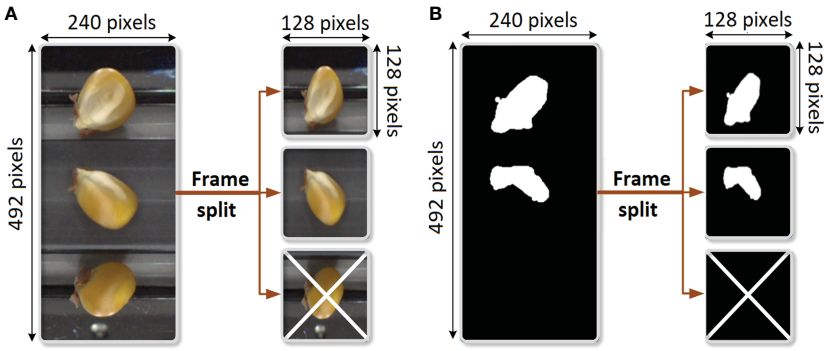


FIGURE 3
Preprocessing for segmentation consists of splitting each dataset image into three 128 x 128 images. Completely negative masks were excluded. **(A)** Preprocessing of RGB image. **(B)** Preprocessing of the mask.

maximize agreement, a loss function using categorical cross-entropy to maximize agreement with positive samples, commonly used in self-supervised learning (Song and Ermon, 2020). To evaluate the performance of the pretrained models, a linear probe — separate from the non-linear projection head included in both models — was attached directly to the encoder and was weight-updated at each step. The backbone and probe were then extracted to calculate validation accuracy for model selection.

2.2.2 Contrastive data augmentation

In many supervised image processing and computer vision tasks, data augmentation is used for the dual purposes of increasing the size of a labeled dataset through synthetic means and improving the diversity of a dataset. For purely supervised purposes, data augmentation can synthetically multiply the dataset's size by altering existing data and increasing the diversity of data to generalize the training set better (Wang and Perez, 2017). Contrastive learning uses heavier image augmentations than would normally be supplied to purely supervised training (Xie et al., 2020). This is due to the reliance of contrastive learning on using augmentations as a model for learning invariance to "style" changes, while the "content" component of a representation remains invariant (Doersch et al., 2015). Thus, heavy stylistic changes should generally benefit the learned representations.

The data augmentations used for our pretraining process were derived from the recommended augmentations particular to SimCLR, consisting of random zoom, random flip, color jitter, and Gaussian noise. NNCLR is less dependent in its performance than SimCLR on the precise type and magnitude of data augmentations used in training; indeed, upon applying augmentations to NNCLR pretraining similar to the full set recommended for SimCLR produced only a 1.6% performance improvement when compared to using only random crop (Dwibedi et al., 2021).

2.2.3 Pretraining setup

Hyperparameter sweeping during pretraining consisted of the variation of the contrastive learning rate, the type of weight initialization applied to the ResNet50 backbone, and data augmentation strength. The learning rate was chosen between 1e-3 and 1e-4, coupling the contrastive learning rate with the classification learning rate of the linear probe. Weight initialization was chosen between ImageNet and random initialization. The data augmentation strength of each augmentation was varied together and explained below. Thus, eight runs were processed for each sweep, and each sweep was repeated three times to ensure precision.

2.3 Classification

2.3.1 Data split

Of the 2697 images manually classified from the unlabeled dataset, there were 1300 oriented images and 1367 non-oriented images. Of the labeled images, 1697 were used for training, with an 800:897 class split in favor of non-oriented images. The rest were divided between validation and testing and were split evenly between the classes. So, 500 images were allocated to each set, with 250 images in each class. During pretraining, the images allocated to the validation and testing were separated from the unlabeled dataset used for contrastive learning, while the labeled training dataset was included, such that 43,286 out of the 44286 total images were used for unlabeled contrastive learning.

2.3.2 Training setup

The training process was set up to facilitate comparison between different models after undergoing end-to-end finetuning. Only ResNet50 was used for the backbones, as is standard in self-supervised model evaluation and as was used in both the NNCLR and SimCLR original papers (Chakraborty et al., 2020; Dwibedi et al., 2021; Shafiq and Gu, 2022). Two backbones for the end-to-end process were chosen from a pretraining sweep with the mentioned self-supervised contrastive architectures, and one backbone was initialized with ImageNet weights.

Data augmentation strength was defined separately for each particular augmentation depending on its configuration specifics: Random zoom acted by cropping to a single rectangle with its shape uniformly chosen between a maximum area of the initial 128x128 2D image shape and a minimum area of either 25% or 75% of the maximum area. Brightness and color transform was accomplished first by taking an identity matrix multiplied by the chosen brightness factor, then adding a matrix with uniformly chosen values selected between the jitter factor and its negative, and secondly by multiplying the original dataset image by this matrix. Brightness jitter increased the brightness of the image by either 50% or 75%, and the jitter factor was either 0.3 or 0.45. Gaussian noise was applied with a standard deviation of either 0.1 or 1.5. The only augmentation kept constant was random flip, constantly at 50% activation chance. Upon evaluation, the two chosen models from this pretraining sweep process—corresponding to the top-left-most two light-green boxes in Figure S2—were backbones pretrained by NNCLR with random initialization at LR = $1e^{-3}$ and SimCLR with ImageNet initialization at LR = $1e^{-3}$.

2.3.3 Feature extraction and finetuning

During training, separate trials were performed for each proportion of annotated data used in classification (1%, 10%, 25%, 100%). As in pretraining, each trial was repeated three times. With 1% and 10% data, a batch size of 4 was used; for 25% data, a batch size of 32 was used; and for 100% data, a batch size of 128 was used. During feature extraction, first, the ResNet-50 backbone from each initialization method was frozen to weight updates, upon which a trainable one-node classifier was constructed with sigmoid activation. Each classifier in every trial was trained for 300 epochs. In finetuning, the backbone was unfrozen, and the entire model was trained for 400 epochs. The same learning rate schedule was used in both phases at the fixed schedule of a 0.5 multiplier every 50 epochs. This process is illustrated in Figure 4.

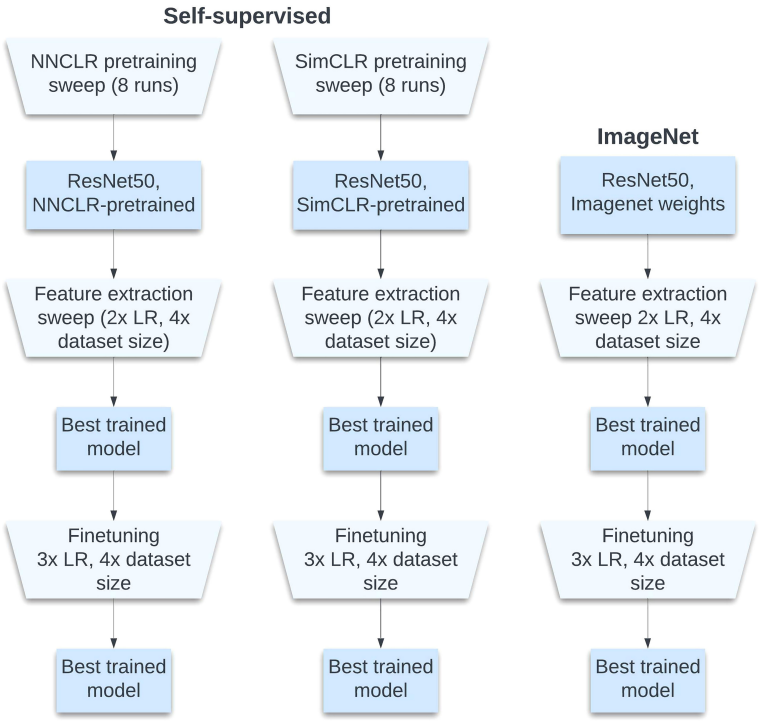


FIGURE 4
Training process. With each sweep over hyperparameters, the best model is chosen for the next round.

2.4 Segmentation

Semantic segmentation is a pixel-level classification problem where the goal is to assign a class label to each pixel of the image. Semantic segmentation of the classified images with the model created above is its natural downstream application. In doing so, full utilization of the QSorter pipeline can be achieved, where along with the immediate results of seed embryo pixel identification, these results can be combined with hyperspectral imaging data in a simple regression problem to pair results in segmentation with results in direct imaging.

2.4.1 Evaluation metrics

The Sørensen–Dice coefficient, also known as the Dice Similarity Coefficient (DSC), is a metric often used in segmentation tasks to evaluate the spatial overlap between two image masks (Taha and Hanbury, 2015). It is given by the equation below:

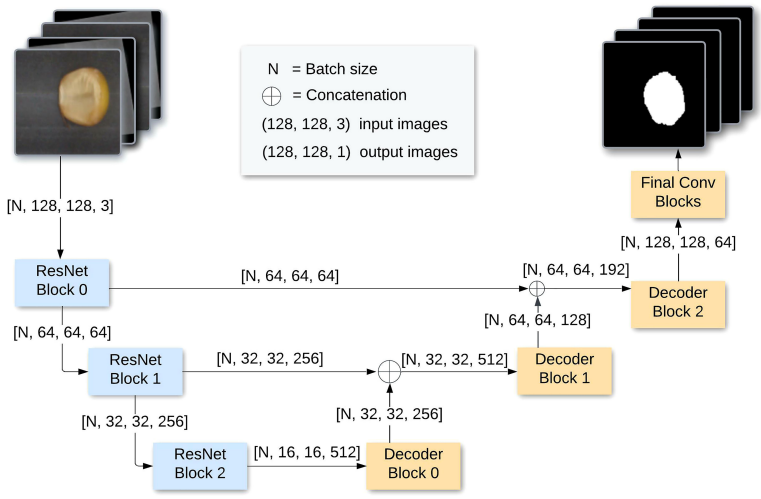


FIGURE 5
U-Net with ResNet50 backbone using filters of channel dimensions [64, 128, 256, 512].

$$Dice(\overrightarrow{y_1}, \overrightarrow{y_2}) = \frac{2\overrightarrow{y_1} \cdot \overrightarrow{y_2}}{\overrightarrow{y_1} + \overrightarrow{y_2}} \quad (1)$$

Here, $\overrightarrow{y_1}$ and $\overrightarrow{y_2}$ are the mask tensors flattened to one dimension. In statistical validation for computer vision tasks, DSC is often preferred over the pixel accuracy metric because DSC ignores true negatives, and pixel classes are often heavily biased toward the (negative) background, especially in binary semantic segmentation.

2.4.2 Model details

U-Net is a convolutional neural network commonly used for semantic segmentation tasks (Zunair and Hamza, 2021). It consists of a symmetric encoder-decoder pair, where the encoder down-samples while increasing the number of channels until a bottleneck tensor, from which the decoder up-samples while reducing the number of channels. For the segmentation task, we used U-Net with ResNet50 used as the encoder to both utilize and compare the self-supervised weights learned during the classification phase, as has been implemented in the literature to considerable advantage (Siddique et al., 2021). In this architecture, the encoder and decoder are not symmetric, as opposed to standard U-Net without a backbone, but skip connections are still fully implemented by limiting the depth of the encoder. Figure 5 shows a U-Net with a ResNet50 as its encoder and four sets of multi-channel feature maps.

2.4.3 Data augmentation

Data augmentation was applied to each training batch to increase the set of distinct training images and to reduce overfitting. Augmentations were coupled between any RGB image and its mask. All augmentations were executed with a 50% application chance. These consisted of combinations of the following: 1) horizontal flip across the vertical middle axis, 2) paired brightness and contrast transform with an application factor uniformly selected from [-0.2, 0.2], and 3) paired scaling and shearing affine transform, the scaling factor uniformly selected from [0.75, 1] and the shear angle uniformly selected from [-π/6, π/6]. Figure S3 shows an example of an augmented image-mask pair.

2.4.4 Training process

Due to the smaller size of the segmentation dataset compared to the classification dataset, ten-fold cross-validation was performed. Using ten folds, ten models were created separately for each backbone and each set of hyperparameters, repeated for each of the three weight initialization types, each trained on a train/validation split of 288/32. With every ten folds, the highest average Dice score across all ten was collected. A model with this set of best-performing hyperparameters was trained on all training data without a validation set for 300 epochs. This model was then evaluated on the full test set. Figure S4 illustrates the cross-validation process. Training and experiments were completed using Google Colab with NVIDIA Tesla T4 and K80 GPUs on 32 GB RAM.

3 Results and discussion

3.1 Classification results

3.1.1 Feature extraction evaluation

We first illustrate the impact of SSL pretraining on annotation efficiency, especially when compared with standard supervised approaches. Figure 6 compares the results of the classifier at various % of training data using a standard supervised loss vs both SimCLR and NNCLR. After feature extraction, (before end-to-end finetuning), both SimCLR and NNCLR were more annotation-efficient and performed better than purely transfer learning-based methods. Listing test results from greatest to least utilization of total available annotated data, the NNCLR-pretrained model had accuracies of 83.6%, 83.2%, 82.0%, and 76.2%; the SimCLR-pretrained model had accuracies of 83.0%, 82.6%, 81.4%, and 78.6%; and the ImageNet-initialized model had accuracies of 82.6%, 81.2%, 77.2%, and 74.6%. At every annotation percentage, the self-supervised models outperformed the ImageNet-based model, with the largest difference at 10% annotation, where the NNCLR-pretrained model outperformed the ImageNet-based model by 4.8%.

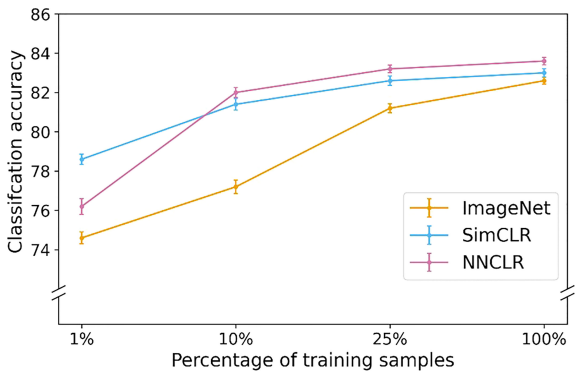


FIGURE 6
Feature extraction classification accuracy versus percentage of training samples for three types of weight initializations.

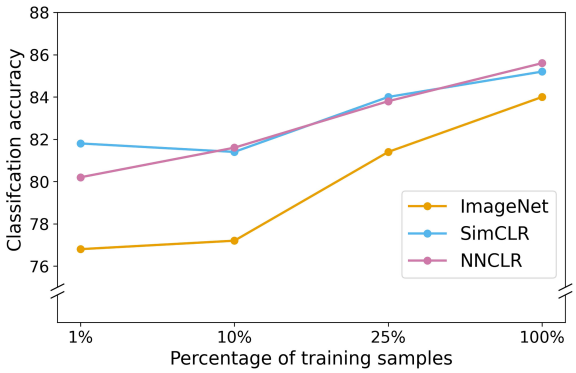


FIGURE 7
Classification accuracy versus percentage of training samples for three types of weight initializations.

3.1.2 Finetuning evaluation

After end-to-end finetuning, both SimCLR and NNCLR were more annotation-efficient and performed better than purely transfer learning-based methods, as shown in Figure 7. Listing test results from greatest to least utilization of total available annotated data, the NNCLR-pretrained model had accuracies of 85.6%, 83.8%, 81.6%, and 80.2%; the SimCLR-pretrained model had accuracies of 85.2%, 84.0%, 81.4%, and 81.8%; and the ImageNet-initialized model had accuracies of 84.0%, 81.4%, 77.2%, and 76.8%. At every annotation percentage, the self-supervised models outperformed all other models, with the largest difference at 1% annotation, where the SimCLR-pretrained model outperformed the ImageNet-based model by 5.0%. Furthermore, at just 1% annotation, SimCLR out-performs the ImageNet-initialized model at 25% annotation. At just 10% annotation, NNCLR also out-performs the ImageNet-initialized model at 25% annotation. We remind the reader that the total available annotated data is only around 5% of the total data (2697 annotated images out of 44,286 total images). SSL pretraining provides a significant boost in model performance, especially at very low total annotated data availability; for instance, a 10% usage of annotated data represents just 270 annotated images!

3.2 Comparisons

Models pretrained with contrastive SSL outperformed transfer learning models in every trial and between all data splits. Table 1 shows the performances of each model compared to the ImageNet-pretrained model. The results of the SimCLR and NNCLR pretrained models outperforming the transfer learning model and being more annotation efficient are clear. The performances of NNCLR and SimCLR were similar to each other among the four annotation percentages, but in training on the full dataset, NNCLR performed slightly better, while SimCLR was more efficient at the lowest data split.

3.3 Segmentation results

Figure 8 shows the test dataset evaluation results after the best models were selected and then finetuned, according to data from the previous three tables. It also shows the validation statistics and hyperparameter set for the chosen model. In Supplementary Information, Table S1 shows the averaged results from 10-fold cross-validation on U-Net with a ResNet50 backbone from weights pretrained with SimCLR, pretrained with NNCLR, and pretrained from ImageNet. Table S2 shows the selected models' hyperparameter set. The U-Net with a SimCLR-pretrained backbone trained at 1e-04 LR and four encoder-decoder filters performed best, with a test DICE score of 0.81 compared to an ImageNet-pretrained backbone at 0.78 DICE score.

The results from this section have a twofold implication: 1) they show U-Net with a backbone loaded with self-supervised pretrained weights can perform well, producing ~0.81 Dice score, and 2) they show semantic segmentation with these backbones outperform those pre-trained with ImageNet. Figure 9 displays three representative results from the segmentation model, including the predicted mask, the true mask, and the input RGB image.

TABLE 1 Relative performance by the accuracy of SimCLR-pretrained and NNCLR-pretrained models as compared to ImageNet preloaded model.

Pretraining	Percentage annotated data used			
	1%	10%	25%	100%
SimCLR	+3.4%	+4.2%	+2.6%	+1.2%
NNCLR	+5.0%	+4.4%	+2.4%	+1.4%

Each entry represents a performance gap in Figure 9. Available annotated dataset size is 2697.

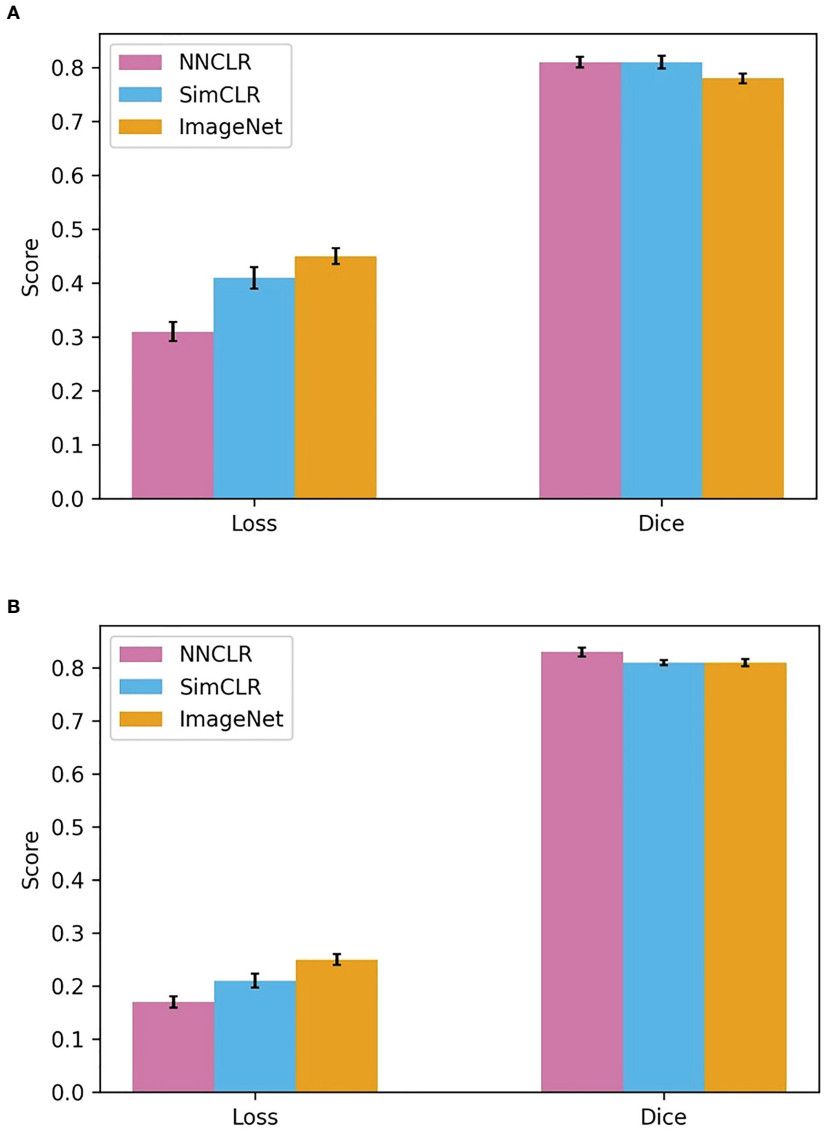


FIGURE 8
Loss and Dice scores for best hyperparameter sets for each weight initialization type. **(A)** Test set results, **(B)** Validation set results.

3.4 Advantages and limitations

In Section 3.1.1, we showed that a SimCLR-pretrained classifier that has gone through end-to-end finetuning out-performs an ImageNet-initialized classifier which uses 96% more annotated training data – the 1% annotation used by a SimCLR-pretrained model resulting in higher accuracy than the 25% annotation used by an ImageNet-initialized model. This is a clear example of the advantage of self-supervised contrastive methods in terms of both human-annotated data efficiency and accuracy. Not only does this curtail the time, labor, and resource-intensive process of annotation as described in the Introduction, but several other by-products of human annotation. For instance, label noise, data bias, the need for domain experts, and imperfect datasets in general are often inevitable with the use of large amounts of annotated data.

Other self-supervised methods have also been developed for computer vision tasks. Our experiments with non-contrastive methods such as SimSiam (Chen and He, 2021) turned out to be examples of the well-known faults of model collapse in non-contrastive self-supervised methods, with models consistently predicting uniform classes, reaching binary classification accuracies of no greater than 55%. We suggest that non-contrastive methods are particularly susceptible to collapse when applied to datasets with relatively homogenous feature spaces such as the applied corn kernel dataset. Furthermore, methods like inpainting (Pathak et al., 2016) have been shown to have poor performance in many applications compared to image augmentation-based methods. Thus, contrastive self-supervised methods which use pretext tasks similar to those of the strong augmentations we applied are particularly suited for processing plant datasets of little species or orientation variation.

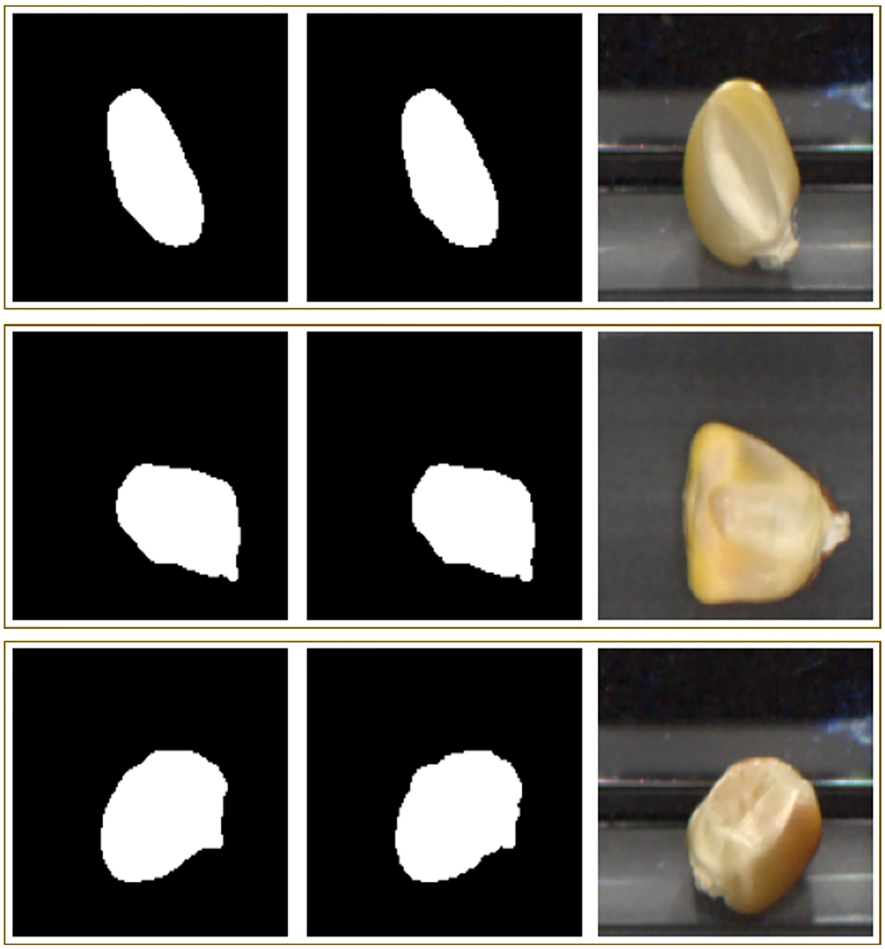


FIGURE 9
Three representative rows of segmentation inputs and outputs. The first column shows the predicted mask, the second shows the true mask, and the third shows the RGB input image.

Although we have found improved performance in applying self-supervised pretraining with all tasks, we expect monotone improvement in fine-tuned performance for classification and segmentation by increasing the size of unlabeled dataset. The clear advantage in relying on pretrained models is that procuring such data is far easier than with similar amounts of labeled data, as would be needed to improve purely supervised classification accuracy. Finally, we expect that such methods to be easily applied to, and very useful to a broad range of plant phenotyping applications. Recent examples of successful applications of such SSL training strategies include disease classification (Nagasubramanian et al., 2022) and insect detection (Kar et al., 2021).

4 Conclusion

From training contrastive learning models and comparing them with purely supervised and transfer learning methods, we found that self-supervised learning produces successful representations of an agricultural dataset applicable for downstream applications. We showed that NNCLR and SimCLR methods performed significantly better than their supervised counterparts, especially for the classification problem. These results also support the usage of strong augmentations in contrastive learning—far stronger than in end-to-end finetuning. In segmentation, self-supervised methods significantly improved over ImageNet pretraining, resulting in accurate masking capabilities and relative embryo size calculation. The combined results further show the transferable nature of self-supervised training. In particular, we illustrated that a single SSL-pretrained model (ResNet50 backbone) could be finetuned and used for two distinct downstream tasks – classification and segmentation. Furthermore, SSL pretraining allowed us to train models with very competitive performance even with very low amounts of total annotated data, for instance, with less than 1% (~400 out of 44000 total images) of annotation. Thus, we have demonstrated that self-supervised learning provides a meaningful path forward in advancing agricultural efficiency with computer vision and machine learning.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://zenodo.org/record/7577017, https://github.com/ddavidd23/ssl_corn.git.

Author contributions

UF, TJ, TL, and BG conceived the project. RW, UF, and TJ conducted physical experiments, data collection and data curation. RW and UF annotated the ground truth images. DD, KN, and TJ developed the machine learning framework. DD performed computational experiments, trained models, and analyzed results. DD wrote the manuscript draft with supervision from BG. All authors revised and edited the manuscript.

Funding

This work was partially supported by AI Institute for Resilient Agriculture (USDA-NIFA #2021-67021-35329), COALESCE: COntext Aware LEarning for Sustainable CybEr-Agricultural Systems (CPS Frontier # 1954556). BG and TL acknowledge support from PSI faculty fellowship.

Acknowledgments

We thank Dr. Candice Gardner at the USDA-ARS Plant Introduction Station.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1108355/full#supplementary-material

References

Aboobucker, S. I., Jubery, T. Z., Frei, U. K., Chen, Y. R., Foster, T., Ganapathysubramanian, B., et al. (2022). "Protocols for in vivo doubled haploid (DH) technology in maize breeding: From haploid inducer development to haploid genome doubling," in *Plant gametogenesis* (Humana, New York, NY: Clifton, NJ), 213–235.

Armstrong, P. R., Tallada, J. G., Hurburgh, C., Hildebrand, D. F., and Specht, J. E. (2011). Development of single-seed near-infrared spectroscopic predictions of corn and soybean constituents using bulk reference values and mean spectra. *Trans. ASABE* 54, 1529–1535. doi: 10.13031/2013.39012

Boote, B. W., Freppon, D. J., de la Fuente, G. N., Lübberstedt, T., Nikolau, B. J., and Smith, E. A. (2016). Haploid differentiation in maize kernels based on fluorescence imaging. *Plant Breed.* 135, 439–445. doi: 10.1111/pbr.12382

Chaikam, V., Molenaar, W., Melchinger, A. E., and Boddupalli, P. M. (2019). Doubled haploid technology for line development in maize: technical advances and prospects. *Theor. Appl. Genet.* 132, 3227–3243. doi: 10.1007/s00122-019-03433-x

Chakraborty, S., Gosthipaty, A. R., and Paul, S. (2020). "G-SimCLR: Self-supervised contrastive learning with guided projection *via* pseudo labelling," in *2020 International Conference on Data Mining Workshops (ICDMW)*. 912–916 (Sorrento, Italy: IEEE).

Chen, X., and He, K. (2021). "Exploring simple Siamese representation learning," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 15745–15753.

Dobrescu, A., Valerio Giuffrida, M., and Tsaftaris, S. A. (2019). "Understanding deep neural networks for regression in leaf counting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA. 4321–4329.

Doersch, C., Gupta, A., and Efros, A. A. (2015). "Unsupervised visual representation learning by context prediction," in *Proceedings of the IEEE International Conference on Computer Vision*. 1422–1430.

Dwibedi, D., Aytar, Y., Tompson, J., Sermanet, P., and Zisserman, A. (2021). "With a little help from my friends: Nearest-neighbor contrastive learning of visual representations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9588–9597.

Fahlgren, N., Gehan, M. A., and Baxter, I. (2015). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* 24, 93–99. doi: 10.1016/j.pbi.2015.02.006

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., et al. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* 2019, 1525874. doi: 10.34133/2019/1525874

Güldenring, R., and Nalpantidis, L. (2021). Self-supervised contrastive learning on agricultural images. *Comput. Electron. Agric.* 191, 106510. doi: 10.1016/j.compag.2021.106510

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., et al. (2021). UAS-based plant phenotyping for research and breeding applications. *Plant Phenomics*, 2021:9840192. doi: 10.34133/2021/9840192

Gustin, J. L., Frei, U. K., Baier, J., Armstrong, P., Lübberstedt, T., and Settles, A. M. (2020). Classification approaches for sorting maize (Zea mays subsp. mays) haploids using single-kernel near-infrared spectroscopy. *Plant Breed.* 139, 1103–1112. doi: 10.1111/pbr.12857

Halcro, K., McNabb, K., Lockinger, A., Socquet-Juglard, D., Bett, K. E., and Noble, S. D. (2020). The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. *Plant Methods* 16, 1–13. doi: 10.1186/s13007-020-00591-8

Jiang, Y., and Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: a review. *Plant Phenomics* 2020, 4152816. doi: 10.34133/2020/4152816

Jiang, H. Y., Zhu, Y. J., Wei, L. M., Dai, J. R., Song, T. M., Yan, Y. L., et al. (2007). Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (Zea mays l.). *Plant Breed.* 126, 492–497. doi: 10.1111/j.1439-0523.2007.01338.x

Jing, L., and Tian, Y. (2020). "Self-supervised visual feature learning with deep neural networks: A survey," in *IEEE transactions on pattern analysis and machine intelligence*, Vol. 43. 4037–4058.

Jones, R. W., Reinot, T., Frei, U. K., Tseng, Y., Lübberstedt, T., and McClelland, J. F. (2012). Selection of haploid maize kernels from hybrid kernels for plant breeding using near-infrared spectroscopy and SIMCA analysis. *Appl. Spectrosc.* 66, 447–450. doi: 10.1366/11-06426

Jubery, T. Z., Carley, C. N., Singh, A., Sarkar, S., Ganapathysubramanian, B., and Singh, A. K. (2021). Using machine learning to develop a fully automated soybean nodule acquisition pipeline (snap). *Plant Phenomics* 2021, 9834746. doi: 10.34133/2021/9834746

Kar, S., Nagasubramanian, K., Elango, D., Nair, A., Mueller, D. S., O'Neal, M. E., et al. (2021). "November. self-supervised learning improves agricultural pest classification," in *AI for Agriculture and Food Systems*.

Kattenborn, T., Leitloff, J., Schiefer, F., and Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS J. Photogrammetry Remote Sens.* 173, 24–49. doi: 10.1016/j.isprsjprs.2020.12.010

Körschens, M., Bodesheim, P., Römermann, C., Bucher, S. F., Migliavacca, M., Ulrich, J., et al. (2021). "Weakly supervised segmentation pretraining for plant cover prediction," in *DAGM German Conference on Pattern Recognition*, Cham. 589–603 (Bonn, Germany: Springer).

Lee, U., Kwon, H., Wu, M., and Wang, M. (2021). Retrospective analysis of the US corn ethanol industry for 2005–2019: implications for greenhouse gas emission reductions. *Biofuels bioprod. Biorefining* 15, 1318–1331. doi: 10.1002/bbb.2225

Lin, X., Li, C. T., Adams, S., Kouzani, A. Z., Jiang, R., He, L., et al. (2023). Self-supervised leaf segmentation under complex lighting conditions. *Pattern Recognition* 135, 109021. doi: 10.1016/j.patcog.2022.109021

Liu, Q., Wang, Z., Long, Y., Zhang, C., Fan, S., and Huang, W. (2022). Variety classification of coated maize seeds based on raman hyperspectral imaging. *Spectrochim. Acta - A: Mol. Biomol. Spectrosc.* 270, 120772. doi: 10.1016/j.saa.2021.120772

Marin Zapata, P. A., Roth, S., Schmutzler, D., Wolf, T., Manesso, E., and Clevert, D. A. (2021). Self-supervised feature extraction from image time series in plant phenotyping using triplet networks. *Bioinformatics* 37, 861–867. doi: 10.1093/bioinformatics/btaa905

McClure, W. F. (2003). 204 years of near infrared technology: 1800–2003. *J. Near Infrared Spec.* 11, 487–518. doi: 10.1255/jnirs.399

Melchinger, A. E., Munder, S., Mauch, F. J., Mirdita, V., Böhm, J., and Mueller, J. (2017). High-throughput platform for automated sorting and selection of single seeds based on time-domain nuclear magnetic resonance (TD-NMR) measurement of oil content. *Biosyst. Eng.* 164, 213–220. doi: 10.1016/j.biosystemseng.2017.10.011

Misra, I., and Maaten, L. V. D. (2020). "Self-supervised learning of pretext-invariant representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6707–6717.

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 419. doi: 10.3389/fpls.2016.01419

Nagasubramanian, K., Jubery, T., Fotouhi Ardakani, F., Mirnezami, S. V., Singh, A. K., Singh, A., et al. (2021). How useful is active learning for image-based plant phenotyping? *Plant Phenome J.* 4, e20020. doi: 10.1002/ppj2.20020

Nagasubramanian, K., Singh, A., Singh, A., Sarkar, S., and Ganapathysubramanian, B. (2022). Plant phenotyping with limited annotation: Doing more with less. *Plant Phenome J.* 5, e20051. doi: 10.1002/ppj2.20051

Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., et al. (2017). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods* 13, 1–12. doi: 10.1186/s13007-017-0173-7

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., and Efros, A. A. (2016). "Context encoders: Feature learning by inpainting," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA 2536–2544.

Paulsen, M. R., and Hill, L. D. (1985). Corn quality factors affecting dry milling performance. *J. Agric. Eng. Res.* 31, 255–263. doi: 10.1016/0021-8634(85)90092-7

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., and French, A. P. (2017). "Deep learning for multi-task plant phenotyping," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*. Venice, Italy 2055–2063.

QualySense *QualySense*. Available at: https://qualysense.com/.

Satake-USA *Optical sorting & processing - rice, wheat & grains - satake USA*. Available at: https://satake-usa.com/ (Accessed 22 Nov. 2022).

Shafiq, M., and Gu, Z. (2022). Deep residual learning for image recognition: A survey. *Appl. Sci.* 12, 8972. doi: 10.3390/app12188972

Siddique, N., Paheding, S., Elkin, C. P., and Devabhaktuni, V. (2021). U-Net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* 9, 82031–82057. doi: 10.1109/ACCESS.2021.3086020

Silvela, L., Rodgers, R., Barrera, A., and Alexander, D. E. (1989). Effect of selection intensity and population size on percent oil in maize, zea mays l. *Theor. Appl. Genet.* 78, 298–304. doi: 10.1007/BF00288815

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Song, J., and Ermon, S. (2020). Multi-label contrastive predictive coding. *Adv. Neural Inf. Process. Syst.* 33, 8161–8173.

Spielbauer, G., Armstrong, P., Baier, J. W., Allen, W. B., Richardson, K., Shen, B., et al. (2009). High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. *Cereal Chem.* 86, 556–564. doi: 10.1094/CCHEM-86-5-0556

Taha, A. A., and Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging* 15, 1–28. doi: 10.1186/s12880-015-0068-x

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* 8, 1190. doi: 10.3389/fpls.2017.01190

Wang, J., and Perez, L. (2017). The effectiveness of data augmentation in image classification using deep learning. *Convolutional Neural Networks Vis. Recognit.* 11, 1–8.

Wang, A., Zhang, W., and Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* 158, 226–240. doi: 10.1016/j.compag.2019.02.005

Ward, O. P., and Singh, A. (2002). Bioethanol technology: developments and perspectives. *Adv. Appl. Microbiol.* 51, 53–80. doi: 10.1016/S0065-2164(02)51001-7

Weinstock, B. A., Janni, J., Hagen, L., and Wright, S. (2006). Prediction of oil and oleic acid concentrations in individual corn (Zea mays l.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Appl. Spectrosc.* 60, 9–16. doi: 10.1366/000370206775382631

Xie, Q., Dai, Z., Hovy, E., Luong, T., and Le, Q. (2020). Unsupervised data augmentation for consistency training. *Adv. Neural Inf. Process. Syst.* 33, 6256–6268.

Xu, Z., Fan, S., Liu, J., Liu, B., Tao, L., Wu, J., et al. (2019). A calibration transfer optimized single kernel near-infrared spectroscopic method. *Spectrochim. Acta - A: Mol. Biomol. Spectrosc.* 220, 117098. doi: 10.1016/j.saa.2019.05.003

Yang, G., Wang, Q., Liu, C., Wang, X., Fan, S., and Huang, W. (2018). Rapid and visual detection of the main chemical compositions in maize seeds based on raman hyperspectral imaging. *Spectrochim. Acta - A: Mol. Biomol. Spectrosc.* 200, 186–194. doi: 10.1016/j.saa.2018.04.026

Zhai, X., Oliver, A., Kolesnikov, A., and Beyer, L. (2019). "S4l: Self-supervised semi-supervised learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South) 1476–1485.

Zhao, X., Vemulapalli, R., Mansfield, P. A., Gong, B., Green, B., Shapira, L., et al. (2021). "Contrastive learning for label efficient semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10623–10633.

Zunair, H., and Hamza, A. B. (2021). Sharp U-net: depthwise convolutional network for biomedical image segmentation. *Comput. Biol. Med.* 136, 104699. doi: 10.1016/j.compbiomed.2021.104699

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fpls.2023.1108560

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service (USDA),
United States

REVIEWED BY
Chunlei Xia,
Chinese Academy of Sciences (CAS), China
Changji Wen,
Jilin Agricultural University, China
Brandon Weihs,
United States Department of Agriculture
(USDA), United States

*CORRESPONDENCE
Qunfeng Niu
niuqunfeng@gmail.com

SPECIALTY SECTION
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

RECEIVED 26 November 2022
ACCEPTED 13 February 2023
PUBLISHED 17 April 2023

CITATION
Wang L, Jia K, Fu Y, Xu X, Fan L, Wang Q, Zhu W and Niu Q (2023) Overlapped tobacco shred image segmentation and area computation using an improved Mask RCNN network and COT algorithm.
Front. Plant Sci. 14:1108560.
doi: 10.3389/fpls.2023.1108560

COPYRIGHT
© 2023 Wang, Jia, Fu, Xu, Fan, Wang, Zhu and Niu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Overlapped tobacco shred image segmentation and area computation using an improved Mask RCNN network and COT algorithm

Li Wang[1], Kunming Jia[1], Yongmin Fu[2], Xiaoguang Xu[2], Lei Fan[2], Qiao Wang[1,3], Wenkui Zhu[4] and Qunfeng Niu[1*]

[1]School of Electrical Engineering, Henan University of Technology, Zhengzhou, China, [2]Xuchang Cigarette Factory, China Tobacco Henan Industry Co, Ltd, Xuchang, China, [3]Key Laboratory of Grain Information Processing and Control (Henan University of Technology), Ministry of Education, Zhengzhou, China, [4]Zhengzhou Tobacco Research Institute of China National Tobacco Company (CNTC), Zhengzhou, China

Introduction: The classification of the four tobacco shred varieties, tobacco silk, cut stem, expanded tobacco silk, and reconstituted tobacco shred, and the subsequent determination of tobacco shred components, are the primary tasks involved in calculating the tobacco shred blending ratio. The identification accuracy and subsequent component area calculation error directly affect the composition determination and quality of the tobacco shred. However, tiny tobacco shreds have complex physical and morphological characteristics; in particular, there is substantial similarity between the expanded tobacco silk and tobacco silk varieties, and this complicates their classification. There must be a certain amount of overlap and stacking in the distribution of tobacco shreds on the actual tobacco quality inspection line. There are 24 types of overlap alone, not to mention the stacking phenomenon. Self-winding does not make it easier to distinguish such varieties from the overlapped types, posing significant difficulties for machine vision-based tobacco shred classification and component area calculation tasks.

Methods: This study focuses on two significant challenges associated with identifying various types of overlapping tobacco shreds and acquiring overlapping regions to calculate overlapping areas. It develops a new segmentation model for tobacco shred images based on an improved Mask region-based convolutional neural network (RCNN). Mask RCNN is used as the segmentation network's mainframe. Convolutional network and feature pyramid network (FPN) in the backbone are replaced with Densenet121 and U-FPN, respectively. The size and aspect ratios of anchors parameters in region proposal network (RPN) are optimized. An algorithm for the area calculation of the overlapped tobacco shred region (COT) is also proposed, which is applied to overlapped tobacco shred mask images to obtain overlapped regions and calculate the overlapped area.

Results: The experimental results showed that the final segmentation accuracy and recall rates are 89.1% and 73.2%, respectively. The average area detection rate of 24 overlapped tobacco shred samples increases from 81.2% to 90%, achieving high segmentation accuracy and overlapped area calculation accuracy.

Discussion: This study provides a new implementation method for the type identification and component area calculation of overlapped tobacco shreds and a new approach for other similar overlapped image segmentation tasks.

KEYWORDS

overlapped tobacco shred, instance segmentation, area computation, Mask RCNN, Cot

1 Introduction

The implementation guidelines set out in Articles 9 and 10 of the WHO Framework Convention on Tobacco Control (FCTC) require the manufacturers and importers of tobacco products to disclose the contents of tobacco products to government authorities, including the type of tobacco shred and each type blending ratio of tobacco shred. Tobacco manufacturers must also have equipment and methods for detecting and measuring tobacco shred components (Acuña, 2017; Niu et al., 2022). The relative proportions of each tobacco shred type (tobacco silk, cut stem, expanded tobacco silk, and reconstituted tobacco shred) impacts the smoke characteristics, physical indicators, and sensory quality of cigarettes (State Tobacco Monopoly Administration, 2009; Chen et al., 2015). Therefore, high-precision and high-efficiency tobacco shred type identification and component determination are crucial to ensuring the quality of the tobacco shred blending process, homogeneity of production, examination of formula design, and accurate identification of tobacco products.

The detection of tobacco shred components has been extensively investigated using both manual and instrumental detection methods. The manual sorting approach involves identifying the tobacco silk, cut stem, expanded tobacco silk, and reconstituted tobacco shred varieties by means of human observation, and then calculating the proportion of various tobacco shreds after weighing. However, this technique has low detection efficiency and is to some extent subjective (Wei et al., 2022). Instrumental detection methods include tobacco shred red, green, and blue (RGB) analysis, hyperspectral imaging analysis, near-infrared spectroscopy, thermal analysis technology, cigarette smoke, anhydrous ethyl ketone, machine vision, and others. Kou et al. (2021) measured the RGB mean value of tobacco powder made from different proportions of the tobacco silk and cut stem varieties. They developed a polynomial regression model that combined the blending ratio and the RGB mean value and proposed a ratio determination method based on RGB image processing to predict cut stem components of tobacco shred. Mei et al. (2021) distinguished the sample components using the spectral data of all sample pixels and proposed a method for identifying tobacco shred components relying on hyperspectral imaging technology. Li et al. (2019) collected the near-infrared spectral data of tobacco shred samples with various component ratios, established an infrared spectral model with the partial least squares regression (PLS) method, and proposed an approach for predicting the blending uniformity of tobacco shreds using infrared spectroscopy. Zhang et al. (2019) utilized the thermogravimetric analysis method to determine the similarity of the tobacco shreds' thermogravimetric reaction rate curves and established a method for calculating the blending uniformity of tobacco shred according to the coefficient of variation between the similarities. Ye et al. (2013) analyzed and compared the smoke indicators and conventional chemical components of tobacco shreds with varying mixing ratios to determine the differences in the blending components. Lin et al. (2020) made use of the fact that the floating rate of expanded tobacco silk under anhydrous ethyl ketone is significantly higher than that of other tobacco shred types, and subsequently developed a method for determining the proportion of components of expanded tobacco silk. Dong et al. (2015); Dong et al. (2016a); Dong et al. (2016b) acquired tobacco shred images through machine vision technology, creating a feature database using different types of tobacco shred images using RGB, HSV (hue, saturation, value) color space pixel variance, contrast, entropy, and others to determine the tobacco shred type. However, all of these detection techniques have some limitations, such as issues with destructive testing, long testing periods, and the incomplete detection of tobacco shred types.

In recent years, machine vision-based deep learning methods have provided advanced and efficient image processing solutions in agriculture. Deep learning methods, combined with machine vision technology, have been widely used in plant disease and pest classification, including the classification of fresh tobacco leaves of various maturity levels (Chen et al., 2021); the classification of tobacco plant diseases (Lin et al., 2022); the classification of wheat spike blast (Fernández-Campos et al., 2021); the classification of rice pests and diseases (Yang et al., 2021); the detection of plant parts such as tobacco leaves and stems (Li et al., 2021); the detection of tomato diseases (Liu et al., 2022); the detection of wheat head diseases (Gong et al., 2020); the detection of brown planthoppers in

rice (He et al., 2020); plant image segmentation, such as tobacco planting areas segmentation (Huang et al., 2021); field-grown wheat spikes segmentation (Tan et al., 2020); rice ear segmentation (Bai-yi et al., 2020; Shao et al., 2021); rice lodging segmentation (Su et al., 2022); photosynthetic and non-photosynthetic vegetation segmentation (He et al., 2022); weed and crop segmentation (Hashemi-Beni et al., 2022); and wheat spike segmentation (Wen et al., 2022). Deep learning methods combined with machine vision technology have been utilized in research focused on the classification of tobacco shred images. Gao et al. (2017) proposed a method for identifying tobacco shreds using convolutional neural networks that is based on differences in the structural characteristics of various tobacco shreds. Zhong et al. (2021) built a recognition model utilizing a residual neural network and optimized the model's pre-training weights, optimization algorithms, and learning rates. They found that both the accuracy and recall rate of the trained model were higher than 96%. Niu et al. (2022) used ResNet50 as the network's primary framework and optimized it by increasing the multi-scale structure, in turn optimizing the number of blocks and loss function. Their experimental results showed that the network's tobacco shred classification accuracy was 96.56%.

In the above research on tobacco shred image classification methods, tobacco shred classification was always carried out using single tobacco shred image samples. In practice, tobacco blends found on the quality inspection line will inevitably contain different types of overlapping and stacked tobacco shreds. The object detection and segmentation methods of overlapping and stacked tobacco shred images have rarely been investigated. However, the type identification and component determination of overlapped and stacked tobacco shred directly affect the calculation accuracy of the blending ratio of tobacco shred components, which is a crucial aspect of research.

The object detection and segmentation methods of overlapped images using machine vision technology have been studied in some fields. Fan et al. (2020) developed a 3D-Mask region-based convolutional neural network (3D-Mask RCNN) for the mass detection and segmentation of overlapping tissue during screening. The 3D-Mask RCNN achieved an average precision (AP) of 0.934 and a false negative rate (FNR) of 0.053. Wang and He (2022) focused on the overlapped images of covered apples in an orchard to perform accurate segmentation. Taking 3D-Mask RCNN as the segmentation network's mainframe, the attention mechanism was added to enhance the network's ability to extract features. The model achieved a recall rate, precision rate, F1 score, and segmentation mean average precision (mAP) of 97.1%, 95.8%, 96.4%, and 0.917, respectively. Yu et al. (2019) achieved accurate segmentation and picking point positioning for overlapping strawberries with a 3D-Mask RCNN and localization algorithm. The average detection precision rate was 95.78%, the recall rate was 95.41%, and the mean intersection over union (MIoU) rate, for instance segmentation, was 89.85%. Qi et al. (2022) concentrated on detecting dense occlusion and overlapped images of auxiliary equipment in an engine room using SsdNet as a mainframe network while adding repulsion loss. The mAP reached 78.95%, which was 5.63% higher than the original SsdNet (SSD). Wen D et al. (2022)'s study set out to identify overlapping bubbles in high void fraction conditions with the use of a convolutional neural network (CNN), and their algorithm reached 85% accuracy under high overlap rate conditions. Wu et al. (2022) employed a residual U-Net network to detect overlapped immunohistochemistry-positive cells in the proposed dataset. Their technique detected 86.04% of the overlapped cells, and the proposed genetic algorithm (GA) outperformed the baseline methods. Su et al. (2019) focused on overlapped ship detection in high-resolution synthetic aperture radar (SAR) imagery using a modified version of the RetinaNet network. The final AP^{50} reached 94.2%. Prasetyo et al. (2020) investigated the performance of two CNN-based segmentation methods, that is, YOLO (you only look once) and Mask RCNN (mask region-based conventional neural network), for separating the heads and tails in images of fish with high variability in terms of their background and illumination, and with overlapping objects. YOLO was high performing, as shown by its 98.6% and 96.73% precision rates. Jia et al. (2020) proposed a model for harvesting robot vision detectors utilizing Mask RCNN to realize the recognition and segmentation of overlapped apples. The precision and recall rates were 97.31% and 95.70%, respectively. Zhang et al. (2022) developed a mask-labeling methodology for particles with a varying degree of overlap that can establish a large and diverse training set without manual labeling. This could be an efficient sample-labeling method.

Regarding the image segmentation of overlapped tobacco shreds, the small size of single-tobacco shreds, their various shapes, and tiny tobacco shreds have complex physical and morphological characteristics, with little difference in macro-scale features between the tobacco silk and expanded tobacco silk varieties, and this complicates the identification and classification of single tobacco shreds with machine vision technology. Furthermore, there are 24 overlapped tobacco shreds derived from four distinct types. The self-winding varieties are more difficult to separate than overlapped types, posing considerable challenges for the segmentation tasks of images of overlapped tobacco shreds and the subsequent calculation of the component area.

This study proposes an overall solution based on an improved Mask RCNN instance segmentation model and an algorithm for the area calculation of overlapped tobacco shred region to identify overlapped tobacco shred types and calculate the area of the overlapped region. The focus is on the identification of overlapped tobacco shred types, as our research object is determining the best method of identifying tobacco shred components for real-world use in field quality inspection lines. This study's contributions to this research area are as follows:

- Establishing two types of original overlapped tobacco shred image datasets, 920 common objects in context (COCO) and 920 visual object classes (VOC). The two datasets consist of images captured from four tobacco shred varieties with 24 overlapped types. The raw overlapped tobacco shred datasets are initially established and applied in a tobacco field, avoiding overfitting and field-specificity.
- Developing an accurate Mask RCNN model to achieve overlapped tobacco shred detection and segmentation utilizing digital images. Segmentation models were

developed and compared using the SsdNet, Deeplap_v3, FcnNet, and RetinaNet architectures with the chosen datasets. The constructed improved Mask RCNN network (Densenet121, U-FPN, anchors parameters) demonstrated the highest instance of segmentation accuracy. It provides good segmentation capability for overlapped tobacco shred images with different sizes and types, outperforming other similar segmentation models.

- Proposing a calculation of overlapped tobacco shred region (COT) algorithm to be first applied to overlapped region identification and overlapped area calculation. This algorithm accurately detects and calculates areas in the images of overlapping tobacco shred, and effectively avoids the negative optimization situation of identifying and calculating overlapped areas.
- Providing a new implementation method for the identification of tobacco shred type and component area calculation of overlapped tobacco shreds and a new approach for other similar overlapped image segmentation tasks.

2 Data acquisition and preprocessing

2.1 Data collection

Cigarette samples in this study were obtained from the Xuchang Tobacco Research Institute of the China National Tobacco Corporation. Each cigarette was a mixture with a certain ratio of four tobacco shred varieties, that is the tobacco silk, cut stem, expanded tobacco silk, and reconstituted tobacco shred varieties (shown in Figure 1). Cigarettes were randomly selected from a specific brand, and thereafter all blended tobacco shreds were obtained. Tobacco shred from a batch with a known serial number was inserted into the vibration device, and then vibration experiments were performed.

The classification results of the vibrated tobacco shred through the vibration experiment are shown in Figure 2. In this experiment, tobacco shred was categorized as (A) single tobacco shred; (B) self-winding tobacco shred; (C) adhesion tobacco shred; (D) inter-overlapped tobacco shred; or (E) stacked tobacco shred. Hence, the overlapped tobacco shreds were defined as one of three overlapped types, namely (A) self-winding tobacco shreds (based on their length); (B) adhesion tobacco shreds (the borders of the two tobacco shreds were connected); or (C) inter-overlapped tobacco shreds, (there was tendency for overlap between two tobacco shreds).

An image acquisition darkroom with four photographic reflectors was designed to obtain high-quality tobacco shred images. Figure 3 is a photograph of an image acquisition system (Niu et al., 2022). The camera and light source were fixed on the bracket, including the universal light source lighting frame, the settings of which could be changed with a fine-tuning knob attached to a 600 mm threaded rod. A Hikvision MV-CE100-30GC industrial camera, a 10-megapixel color camera, was used and equipped with a MVL-HF1224M-10MP 12 mm focal length Hikvision industrial lens. The Hikvision Technology Industrial Ring Angle Light Source R120-80-25 was selected as a ring light source to provide uniform brightness in the shooting field of view, eliminating the influence of shadows cast by the tobacco shreds. The computer and the industrial camera were connected by a network

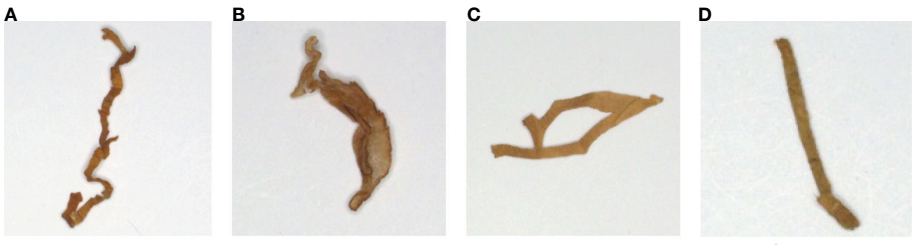


FIGURE 1
The four tobacco shred varieties: **(A)** tobacco silk; **(B)** cut stem; **(C)** expanded tobacco silk; and **(D)** reconstituted tobacco shred.

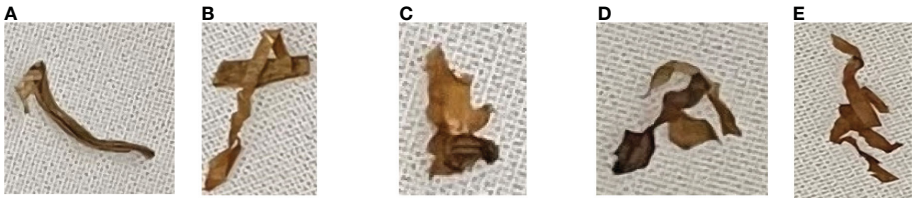


FIGURE 2
The vibrated tobacco shred types: **(A)** single tobacco shred; **(B)** self-winding tobacco shred; **(C)** adhesion tobacco shred; **(D)** inter-overlapped tobacco shred; and **(E)** stacked tobacco shred.

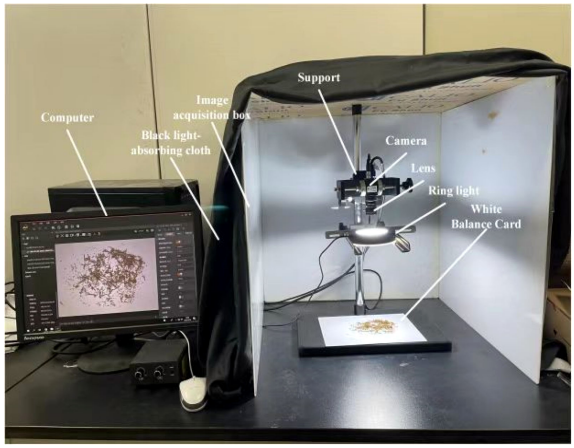


FIGURE 3
The image acquisition system.

cable, ensuring that the tobacco shred images were transmitted over a stable connection at high speed.

Table 1 shows the overlapped tobacco shred types and the number of images utilized in this study. The overlapped tobacco shred images were obtained from the four tobacco shred varieties [cut stem (G), expanded tobacco silk (P), tobacco silk (Y), and reconstituted tobacco shred (Z)]. A total of 920 overlapped tobacco shred images of 24 overlapped tobacco types (i.e., four self-winding tobacco shreds, 10 adhesion tobacco shreds, and 10 inter-overlapped tobacco shreds) were taken. The size of a single image was 3,840 × 2,748 pixels. The number of tobacco shred images taken differs for various overlapped tobacco shreds. For the tobacco shred overlaps of the same type (for example, GG is overlapped with cut stem and cut stem), the ratio of self-winding, to adhesion, to inter-overlapped types was set to 1:1:2. For the overlap of different types of tobacco shred (for example, GP is cut stem overlapped with expanded tobacco silk), the ratio of adhesion to inter-overlapped types was set to 1:1. As sample images could not easily be obtained, and a wide variety of overlapped types exist, the ratio of the training set to the testing set was set to 8:2.

2.2 Data preprocessing

It was observed that the segmentation model's training and testing times increased considerably if the tobacco shred object in

TABLE 1 Overlapped tobacco shred datasets.

Overlapped tobacco shred types	Self-winding tobacco shred	Adhesion tobacco shred	Inter-overlapped tobacco shred	Number of images
GG				98
GP				96
GY				80
GZ				95
PP				95

(Continued)

10.3389/fpls.2023.1108560

TABLE 1 Continued

Overlapped tobacco shred types	Self-winding tobacco shred	Adhesion tobacco shred	Inter-overlapped tobacco shred	Number of images
PY				89
PZ				97
YY				91
YZ				93
ZZ				86
Total				920

GG, cut stem and cut stem; GP, cut stem and expanded tobacco silk; GY, cut stem and tobacco silk; GZ, cut stem and reconstituted tobacco shred; PP, expanded tobacco silk and expanded tobacco silk; PY, expanded tobacco silk and tobacco silk; PZ, expanded tobacco silk and reconstituted tobacco shred; YY, tobacco silk and tobacco silk; YZ, tobacco silk and reconstituted tobacco shred; ZZ, reconstituted tobacco shred and reconstituted tobacco shred.

the overlapped tobacco shred images obtained by the image acquisition system was small, and if the images contained a lot of background information. The image preprocessing approach starts by treating the overlapped tobacco shred images with the OpenCV algorithm, then finding the minimum circumscribed circle of the object and cutting the contour. The algorithm flow chart is shown in Figure 4. The image preprocessing algorithm ensures that the invalid background information in the picture is reduced by preserving the foreground information of the image, thereby significantly reducing the size of the image.

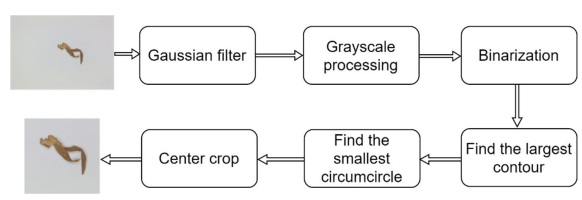


FIGURE 4
Overlapped tobacco shred image preprocessing process.

The preprocessed overlapped tobacco shred images were labeled using LabelMe, an image annotation tool, to generate corresponding mask images. Thereafter, the COCO official datasets were used to develop the code and the overlapped tobacco shred datasets of the COCO data type. The four tobacco shred regions in the image were marked, and the rest were taken as the image's background. The labeled image of the inter-overlapped GP is shown in Figure 5.

3 Methods

3.1 Segmentation method

There are three main challenges in the task of overlapped tobacco segmentation:

(1) The size of a single tobacco shred is small, the shape of the tobacco shred is variable, and tiny tobacco shreds have complex physical and morphological characteristics; in particular, there is substantial similarity between expanded tobacco silk and tobacco silk.

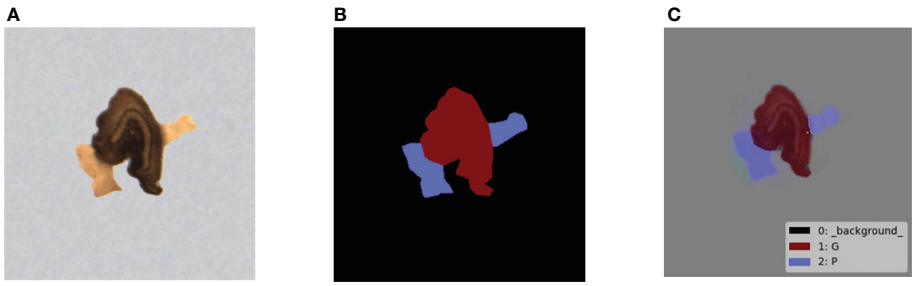


FIGURE 5
Overlapped tobacco shred example of instance segmentation: **(A)** original image; **(B)** mask image of instance segmentation; and **(C)** visualization of the mask image. G, cut stem; P, expanded tobacco silk.

(2) There are 24 types of overlapped tobacco shreds that can be extracted from the four different varieties, and shreds of the self-winding type are not easily distinguished.

(3) Overlapped tobacco shreds are small target objects, meaning that their segmentation is difficult.

The following approaches were used to overcome the above challenges:

- Mask RCNN is a state-of-the-art CNN-based method in which the detection and segmentation of objects are performed simultaneously to address broad problems associated with overlap across multiple domains. Therefore, Mask RCNN, as the mainframe of the segmentation network, was used to complete the target detection of the overlapped tobacco shreds in the image and segment various types of tobacco shreds.
- In the case of complex classification tasks and limited datasets, the Resnet in the backbone of the Mask RCNN was replaced with Densenet121, because the latter adopts a dense connection mode between layers. As a result, multiple rounds of shallow information were used to increase the ability of the Mask RCNN to extract tiny features in the shallow information of the overlapped tobacco shreds. The Densenet121 can achieve better performance using fewer datasets compared with the Resent.
- Because of the rich shallow features in the tiny overlapped tobacco objects, the deep features contain less target information. Accordingly, the feature pyramid network (FPN) in Mask RCNN was changed to a U-FPN. In contrast to the FPN, U-FPN performs feature multiplexing on C2 and C3 and features multiplexing on P2, P3, P4, and P5, which significantly enhances the utilization rate of tiny features in shallow information.
- The anchors parameters in the region proposal network (RPN) were optimized, and the size and aspect ratios suitable for the small objects of overlapped tobacco shred were designed to ensure that the RPN could extract region of interest (ROI) features from different levels efficiently. In this way, the extraction and bounding boxes performance of small objects was significantly improved without a correspondingly large increase in the computational cost, and this increased the model's ability to detect tobacco shred and enhanced its accuracy.

3.2 Segmentation network

3.2.1 Overall model framework of improved Mask RCNN

The Mask RCNN network was introduced by He et al. (2017). The network has achieved excellent results in various tasks by utilizing the Microsoft Common Objects in Context (MS COCO) dataset, including object detection, instance segmentation, and keypoint detection. The mainframe of the segmentation network in this paper adopts the improved Mask RCNN network. The network structure of the improved Mask RCNN network is composed of main modules with backbone (CNNs and FPNs), RPNs, ROI aligns, fully convolutional networks (FCNs), and fully connected (FC) layers.

The overall framework of the improved Mask RCNN network is shown in Figure 6. It can be seen that the model input size is 500 × 500 pixels of overlapped tobacco shred images, and that the backbone network uses the DenseNet121+U-FPN group to perform feature extraction that obtains the feature map. Moreover, the feature map output from the backbone is fed to the RPN to generate proposals. Subsequently, the ROI output from the RPN is mapped to extract the corresponding overlapped tobacco shred features in the shared feature map. Finally, the instance segmentation of overlapped tobacco shred images is completed with FC layers and FCNs. The model's output is the overlapped tobacco shred type.

3.2.2 Improved Mask RCNN performance brought about by a change in backbone

Because of the different morphological characteristics and the slight differences in the features of the overlapped tobacco shreds, extracting the features of different overlapped tobacco shreds is challenging, particularly in the overlapped region. In the Mask RCNN network, using Resnet50 to extract different levels of features from the overlapped tobacco shred input images is less effective. However, an increased number of Densenet121 network layers can enhance the ability to extract small target detail information of

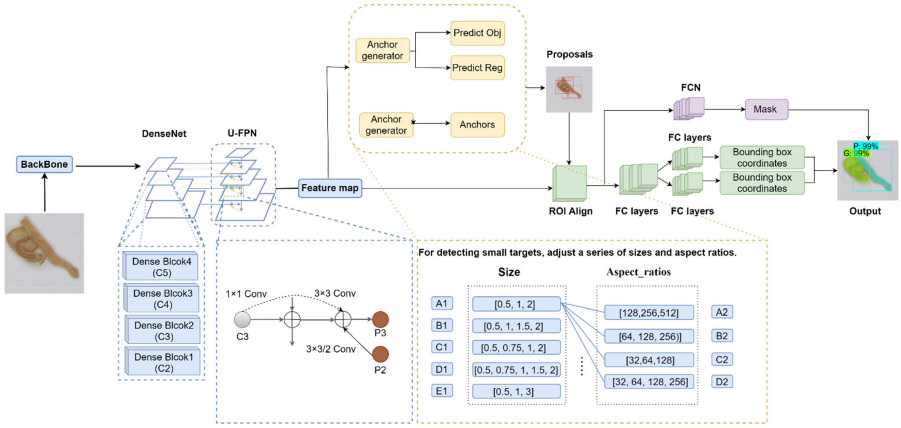


FIGURE 6
The overall framework of the improved Mask RCNN. FC layers, fully connected layers; FCNs, fully convolutional networks; Mask RCNN, mask region-based conventional neural network; ROI align, region of interest align.

different tobacco shreds. The dense connection form can effectively extract the small differences in the features, and features of different tobacco shreds in the overlapped regions. Therefore, DenseNet121 can effectively extract small-sized target detail information and small features in larger-sized overlapped tobacco shreds, which enhances the overall feature extraction ability of the model in regions with overlapped tobacco shred, and to a certain extent solves the problem of shallow feature loss. Densenet121 was set to four feature extraction layers, that is Dense Block 1, Dense Block 2, Dense Block 3, and Dense Block 4. The four feature return values have undergone different downsampling times (i.e., 2, 3, 4 and 5s) and lateral connections. Accordingly, Densenet121 constitutes a new type of backbone, as shown in Figure 7.

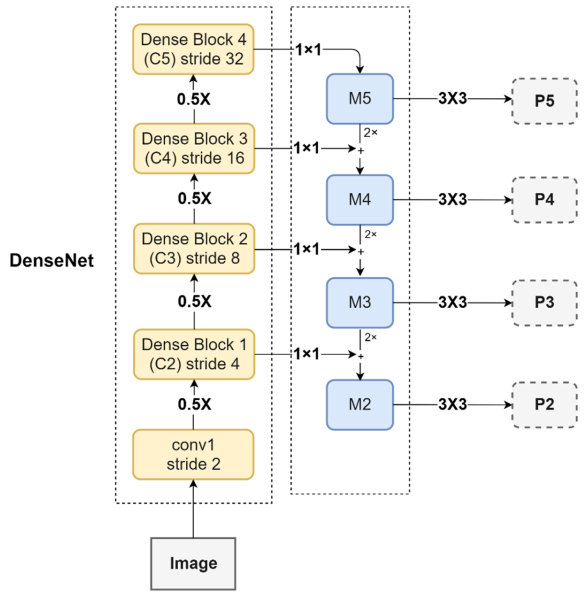


FIGURE 7
A new type of backbone (Densenet121+FPN).

3.2.3 Improved Mask RCNN about U-FPN

After being extracted by the CNN network, the receptive field of the shallow-feature map is small, and more detailed information about the tobacco shred targets is produced, whereas the receptive field of the deep-feature map is large, and the information it produces about these tiny targets is less detailed. Although the top-down and same-layer connected structure of the FPN combines deep and shallow features to a certain extent to meet the needs of subsequent classification and detection of overlapped tobacco shreds, it still cannot make up for the complete usage of small features in the shallow parts. The shallow feature information of small targets overlapped tobacco shred detection is rich and vital. Therefore, a bottom-up and horizontal feature multiplexing structure was added based on the top-down structure, and the shallow information was transmitted to each feature layer (P3, P4, P5, and P6), which enhanced the effective use of shallow feature information. The modified FPN network structure was named U-FPN, as shown in Figure 8. P3 in Figure 8 indicates that adding P2 and C3 to P3 through 3 × 3/2 Conv and 3 × 3Conv helped obtain the shallow information of the P2 layer, and that incorporating the C3 layer into the P3 layer enhanced the fusion of the general shallow feature information.

3.2.4 Improved Mask RCNN about RPN

In the Mask RCNN model, the scales and aspect ratios of the anchor were set to [128, 256, 512] and [1:1, 1:2, 2:1], respectively, with nine reference anchors being set for each position on the feature map. The RPN selects and adjusts anchor output ROIs according to the features of each stage. For instance, in P2, the layer's feature map was 256 × 256, and the step was 4; hence each pixel on P2 generated a 4 × 4 anchor frame with an area of 16 based on the current coordinates. According to the scales and aspect ratios of the anchor, bounding boxes of the three sizes and three shapes were generated at each pixel point. The foreground and background classification and offset regression of the bounding boxes were conducted after two convolution layers.

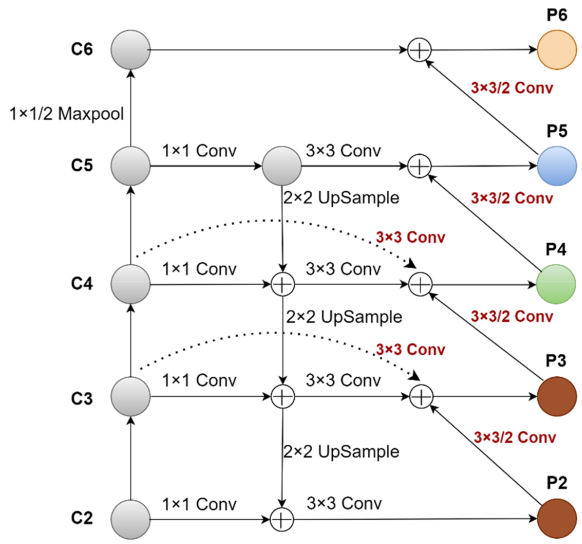


FIGURE 8
The U-FPN network structure.

The minor anchor scale in the Mask RCNN was 128 × 128, but there were many small targets in the overlapped tobacco shred, some of which were much smaller than this scale, resulting in the model's inability to detect this target object. Ideally, the smaller the target, the denser the anchors to cover all the candidate regions, and the larger the target, the fewer and sparser the anchors should be. Otherwise, the anchors overlap and cause redundant computation. However, the anchors parameter set in Mask RCNN, focusing on small targets in the images of overlapped tobacco shred objects, led, to a certain extent, to the anchors for small targets being few and sparse, and the anchors for detecting large targets being many and dense. In this case, the detection performance of small objects can be significantly improved by adjusting appropriate anchor scales and aspect ratios without greatly increasing the amount of computation.

According to the statistics of the aspect ratios in each batch of images and the pixel size for tobacco shred of [0, 0.5, 0.6, 0.8, 1.0, 1.3, 1.5, 2.0], a series of anchor scales and aspect ratios were designed as [128, 256, 512], [64, 128, 256], [32, 64, 128], [32, 64, 256] and [0.5, 1, 2], [0.5, 1, 1.5, 2], [0.5, 0.75, 1, 2], [0.5, 0.75, 1, 1.5, 2], and [0.5, 1, 3]. Finally, the anchor scale and size were experimentally determined as [32, 64, 128, 256] and [0.5, 1, 1.5, 2], respectively.

3.3 Area calculation of overlapped tobacco shred algorithm

Based on the improved Mask RCNN network, the above detection model can effectively achieve instance segmentation for overlapped tobacco shreds with different shapes and forms and obtain the contours for the overlapped tobacco shred targets. Based on this, it is possible to calculate the pixel area and respective area proportions of the corresponding tobacco shreds for the mask image through the OpenCV algorithm. However, the overlapped region of the occluded tobacco shred cannot be obtained. The omission of the overlapped area in the covered tobacco shred directly leads to errors in the calculation of the respective areas for different tobacco shreds and the total area in the tobacco shred group during the subsequent determination of its components.

The area calculation of overlapped tobacco shreds requires using the improved Mask RCNN network to generate a mask image of the overlapped tobacco shred, determine the occluded tobacco shred, draw and fit the overlapped region according to the distribution of the occluded overlapped tobacco shred, and determine the actual overlap region with the fitted overlapped region and the unoccluded area. Finally, the pixel area in the overlapped part is calculated. The algorithm for area calculation in the overlapped region, named the COT algorithm, is as follows, and the specific process is shown in Figure 9:

- Determine the occluded tobacco shred object. First, grayscale and binarize the mask image, and calculate and count the number of tobacco shred contours. A single contour is an unoccluded tobacco shred (see unoccluded image in Figure 9). Multiple contours are occluded tobacco shreds (see occluded image in Figure 9).
- Fit the overlapped region in the tobacco shred (see Mask processing in Figure 9). The multiple contours of the shrouded tobacco shred are cyclically judged, and two contours are found according to the size of the contour area. First, use the

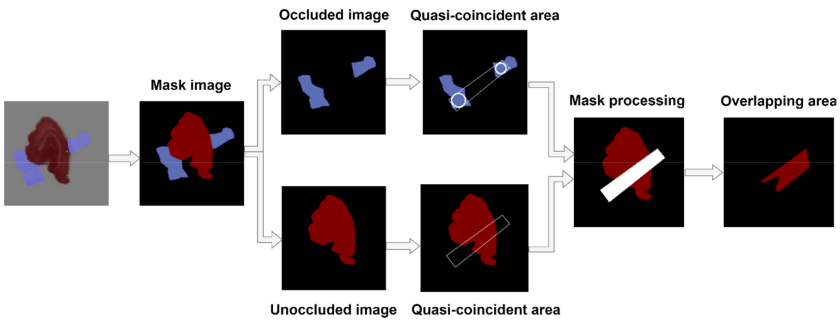


FIGURE 9
Specific process of the calculation of overlapped tobacco shred region (COT) algorithm.

cv2.pointPolygonTest and cv2.minMaxLoc in the OpenCV function to construct the smallest rectangle inside the outline. Then draw the smallest inscribed circle round _1 with the center of the rectangle as the dot, the side length as the diameter, the center as (x1, y1), and the diameter as d1. Then, draw the smallest inscribed circle round _2 of the second contour, whose center is (x2, y2), and whose diameter is d2. Connect the centers (x1, y1) and (x2, y2) of the two inscribed circles, and draw a straight line L0. By extending the straight line L0, draw outer tangent lines L1 and L2 with the diameter of each circle for round _1 and round _2. Draw the fitted overlap region trapezoid _1 according to L1 and L2 (see the quasi-coincident area in Figure 9).

- Determine the overlapped region of the actual tobacco shred and pixel area by calculating the overlapped region. First, use cv2.fillPoly function to generate a mask image for the fitting area, then perform a mask operation according to the outline of the unoccluded image and find the overlapped area (see the overlapped areas in Figure 9). Finally, the pixel area of overlapped area is calculated using OpenCV

3.4 Evaluation index

3.4.1 Improved Mask RCNN

The COCO evaluation index is the current mainstream target detection and instance segmentation evaluation index. This paper uses the COCO evaluation indicators, i.e., training time (T-time) and prediction time (P-time) as evaluation indicators for improved Mask RCNN and other baseline segmentation models. As an image of overlapped tobacco shreds is a small target object, among the COCO evaluation indicators, six indicators (AP,AP^{50},AP^{75},AP^{s}, AR^{10} and AR^{s}) were selected for network performance evaluation, where AP^{50} and AR^{s} represent the precision and recall rates, respectively. The higher these values, the more ideal the segmentation model (Tong et al., 2020).

Average precision (AP):

AP % AP at IOU = 0.50: 0.0.5: 0.95 (primary challenge metric)

AP^{50} % AP at IOU = 0.50 [PASCAL Visual Object Classes (PASCALVOC) metric]

AP^{75} % AP at IOU = 0.75 (strict)

AP Across scales L:

AP^{s} % AP for small objects: area< 32^2

Average recall (AR):

AR^{10} % AR give 10 at detections per image

AR Across scales:

AR^{s} % AR for small objects: area< 32^2

3.4.2 COT algorithm

This paper uses Eqs 1–4 to calculate the actual area ratio (AAR), COT area ratio (CAR), average actual area ratio (Avg_AAR), and average COT area ratio (Avg_CAR) as evaluation indicators of the COT algorithm. The higher the CAR and Avg_CAR values, the higher the COT algorithm's area calculation accuracy to compensate for the overlapped region of occluded tobacco shred and the more ideal the algorithm.

$$AAR = \frac{AD}{Area_1 + Area_2} \tag{1}$$

$$CAR = \frac{AD + CD}{Area_1 + Area_2} \tag{2}$$

$$Avg_AAR = \frac{\sum_{i=1}^{n} AAR_i}{n} \tag{3}$$

$$Avg_CAR = \frac{\sum_{i=1}^{n} CAR_i}{n} \tag{4}$$

AAR actual area ratio
CAR COT area ratio
AD actual detection area
CD detection area
Area_1 the pixel area of tobacco shred 1
Area_2 the pixel area of tobacco shred 2
Avg_AAR average actual area ratio
Avg_CAR average COT area ratio
n total number of AAR or CAR

The overlapped tobacco shred samples consist of four randomly combined tobacco shreds in pairs. Tobacco shred 1 is defined as the unshielded tobacco shred in the overlapped tobacco shred samples, and tobacco shred 2 is the occluded tobacco shred for evaluating the COT algorithm. The sum of the area of tobacco shred 1 and tobacco shred 2 is the arbitrary actual total area of the overlapped tobacco shred. The actual detection (AD) area is the complete pixel area of the overlapped tobacco shred image calculated by the OpenCV algorithm without considering the occlusion of any overlapped tobacco shred sample. The COT detection (CD) overlapped area is the pixel area of the overlapped area calculated with the COT algorithm upon obtaining the overlapped region outline of tobacco shred 2 in the overlapped tobacco shred image.

4 Results

4.1 Implementation details

4.1.1 Experimental platform

The experiment in this paper was performed on a Windows 10 operating system. The GPU used was the GeForce GTX 3080 (10GB video memory), the processor used was the Intel Core i7-12700K CPU@3.61GHz, and the running memory was 64 GB. The model's construction, training, and testing were implemented in Python using the PyTorch deep learning framework. In addition, the

CUDA 11.0 parallel computing framework was used alongside the Pycharm development environment.

4.1.2 Details of segmentation model training

For the overlapped tobacco shred dataset, the training set images were randomly shuffled before input to reduce the influence of the image sequence on the model. During the model training process, the batch size for model training was set to 8, the number of training arguments was taken as 60, and the initial learning rate was taken as 0.08. In the model gradient optimization, the gradient descent was performed after multiple iterations using the stochastic gradient descent (SGD) optimizer and the learning rate was attenuated during the model training process to obtain better segmentation performance.

4.2 Results of the segmentation model

4.2.1 Improved Mask RCNN performance test

The overlapped tobacco shred datasets were used for network training and testing using the improved Mask RCNN. Figure 10 shows the segmentation effect diagrams for 24 overlapped types of tobacco shreds. The proposed method can accurately classify self-winding and adhesion tobacco shreds. The average target recognition accuracy for the four self-winding tobacco shreds was 99% (A99%, J99%, O99%, and V99%). The average object recognition accuracy for overlapped tobacco shreds across nine adhesion types was 94% (B99%, D99%, F99%, H66%, K96%, M99%, P90%, T99%, and W99%). The average target recognition accuracy for the inter-overlapped tobacco shreds that were difficult to segment was 86.9% (C99%, E99%, G41%, I80%, L99%, N67%, Q99%, S99%, and X99%). Therefore, the above analysis indicates that various overlapped tobacco shreds can be accurately identified using the improved Mask RCNN. Although the improved network leads to increased training and inference time, it can still ensure accurate segmentation and recognition relatively quickly.

The experimental results show that after 60 training rounds, the loss and mAP of the model remained stable, the training time was 4454.8 s, and the inference time was 0.04 s. The average precision (AP^{50}) and average recall (AR^{s}) for the object detection performance of the improved Mask RCNN model were 90.2% and 75.2%, respectively. Additionally, the AP^{50} and AR^{s} for the segmentation performance of the Mask RCNN model were 89.1% and 73.2%, respectively. The improved network performance is shown in Table 2. The segmentation model performance is separately explained below.

4.2.2 Performance test of the Densenet121

This section evaluates the model performance of the DenseNet121 for segmenting overlapped tobacco shreds. The following networks were selected as baseline backbones: Vgg11 (Simonyan and Zisserman, 2014), MobileNet (Howard et al., 2017), Resnet50 (He et al., 2016), Resnet101 (Xu et al., 2020), and Densenet121 (Huang et al., 2017). The performance index comparison of different CNN backbone networks is shown in Table 3.

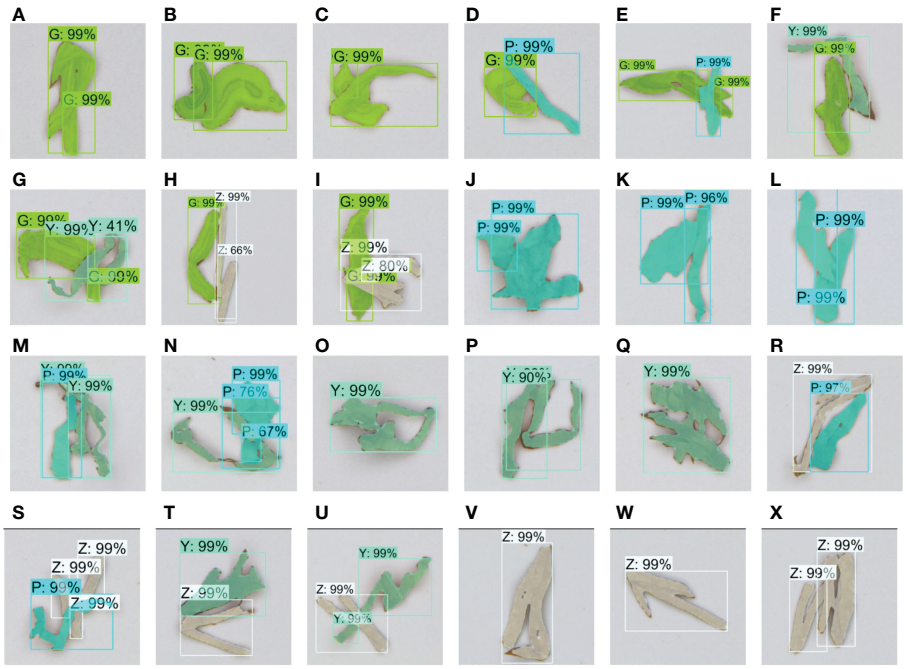


FIGURE 10
A–X are recognition results of 24 overlapped tobacco shreds, including self-winding, adhesion, and inter-overlapped GG, PP, YY, and ZZ, and adhesion and inter-overlapped GP, GY, GZ, PY, ZP, and ZY. G, cut stem; P, expanded tobacco silk; Y, tobacco silk; Z, reconstituted tobacco shred; GG, cut stem and cut stem; GP, cut stem and expanded tobacco silk; GY, cut stem and tobacco silk; GZ, cut stem and reconstituted tobacco shred; PP, expanded tobacco silk and expanded tobacco silk; PY, expanded tobacco silk and tobacco silk; PZ, expanded tobacco silk and reconstituted tobacco shred; YY, tobacco silk and tobacco silk; YZ, tobacco silk and reconstituted tobacco shred; ZZ, reconstituted tobacco shred and reconstituted tobacco shred.

TABLE 2 Improved Mask region-based convolutional neural network (RCNN) performance.

Network structure		AP	AP^{50}	AP^{75}	AP^{s}	AR^{10}	AR^{s}	T-time (s)	P-time (s)
Improved Mask RCNN	**Detection**	**0.677**	**0.902**	**0.787**	**0.749**	**0.752**	**0.752**	**4454.8**	**0.044**
	Segmentation	**0.641**	**0.891**	**0.733**	**0.728**	**0.732**	**0.732**		

Mask RCNN, mask region-based conventional neural network.

TABLE 3 The performance index comparison of different convolutional neural network (CNN) backbone networks.

Network Structure	AP	AP^{50}	AP^{75}	AP^{s}	AR^{10}	AR^{s}	T-time (s)	P-time (s)
Mask RCNN Vgg11	0.523	0.817	0.547	0.646	0.648	0.648	4610.0	0.053
Mask RCNN MobileNet	0.602	0.840	0.661	0.695	0.697	0.697	3986.2	0.016
Mask RCNN ResNet 50	0.466	0.805	0.476	0.582	0.583	0.583	3725.4	0.035
Mask RCNN ResNet 101	0.533	0.834	0.590	0.640	0.642	0.642	4771.3	0.043
Mask RCNN Densenet 121	**0.612**	**0.861**	**0.666**	**0.698**	**0.695**	**0.695**	**4466.9**	**0.041**

Mask RCNN, mask region-based conventional neural network.

Table 3 compares the Mask RCNN using Densenet121 as the backbone with Mask RCNN using lightweight networks such as ResNet50 and MobileNet as the backbone. Although the training time (4466.9s) and inference time (0.041s) of the DenseNet121 were longer than those of the ResNet50 (3725.4s and 0.035s, respectively) and the MobileNet (3986.2s and 0.016s, respectively), DenseNet121 had the best AP^{50} and AR^{s} at 0.861 and 0.695, respectively, and the remaining indices were also the better than or comparable to the other backbones. The training time and inference time of the DenseNet121 backbone network were minimal, and the performance indicators were the best when compared with the Mask RCNN using Vgg11 or ResNet101 as the backbone. This means that in lightweight and deep networks, Mask RCNN with the DenseNet121 backbone network can effectively identify and segment overlapped tobacco shreds.

4.2.3 Performance test based on U-FPN

Based on the improved proof in 3.2.2, the model performance for small-target detection based on the U-FPN algorithm was evaluated. The performance indices of Mask RCNN–FPN, Mask RCNN–U-FPN, and Mask RCNN–U-FPN–Densenet121 are listed in Table 4.

Table 4 indicates that the Mask RCNN–U-FPN network shows noticeable improvements in small-target detection and segmentation performance compared with the Mask RCNN-FPN network, which proves that even when the backbone was the ResNet50 network, U-FPN showed excellent network performance. Compared with the Mask RCNN–U-FPN network, Mask RCNN–U-FPN–Densenet121 demonstrated effective network performance improvements without significantly increasing the network training and prediction time. The AP^{50} and AR^{s} of the Mask RCNN–U-FPN–Densenet121 network were the best recorded, at 0.877 and 0.727, respectively. Except for time, the other indicators were still the best in the Mask RCNN–U-FPN–Densenet121 network, which proves U-FPN's effectiveness.

4.2.4 Performance test based on RPN

In this section, Mask RCNN–U-FPN–Densenet121 is referred to as P-Mask RCNN for convenience. A series of Anchor sizes (A1–E1) and aspect ratios (A2–E2) was designed on the premise that the aspect ratios of all images was [0, 0.5, 0.63, 0.8, 1.0, 1.26, 1.59, 2.0], as shown in Figure 11. Additionally, the anchors parameters in the RPN of the P-Mask RCNN model were adjusted, the overlapped tobacco shred datasets were inputted, and the model's network performance was compared. Table 5 shows different P-Mask RCNN performances under different sizes and aspect ratios. In Table 5, the P-Mask RCNN-A1A2 represents the original network with the default anchors parameters. The networks with adjusted anchors parameters range from P-Mask RCNN-A1B2 to P-Mask RCNN-E1D2.

Table 5 shows that choosing inappropriate anchors parameters, such as B1A2, B1B2, B1C2, E1B2, and E1D2, led to a decline in network performance, and this proves the importance of improving

TABLE 4 The performance index comparison of different feature pyramid networks (FPNs).

Network structure	AP	AP^{50}	AP^{75}	AP^{s}	AR^{10}	AR^{s}	T-time (s)	P-time (s)
Mask RCNN–FPN	0.466	0.805	0.476	0.582	0.583	0.583	3725.4	0.035
Mask RCNN–U-FPN	0.498	0.819	0.493	0.609	0.608	0.608	0.608	0.608
Mask RCNN–U-FPN–Densenet121	**0.637**	**0.877**	**0.689**	**0.724**	**0.727**	**0.727**	**4523.1**	**0.042**

Mask RCNN, mask region-based conventional neural network.

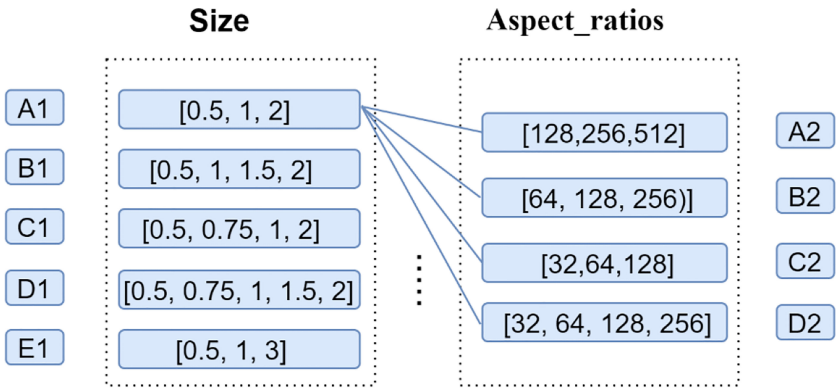


FIGURE 11
A series of Anchor sizes (A1–E1) and aspect ratios (A2–E2).

parameters. In addition, it indicates that changing the anchors parameter can effectively improve the network performance at the cost of a slight increase in training time. Finally, it was determined that the Anchor sizes [0.5, 1, 1.5, 2] and the aspect ratios [32, 64, 128, 256] performed the best. Although the training time and prediction time of the P-Mask RCNN-B1D2 network were slightly higher than with other anchors parameters, performance indices were better.

4.2.5 Comparison with other instance segmentation methods

In order to further prove the effectiveness of the improved Mask RCNN at solving the problem of overlapped tobacco shred image segmentation, SsdNet (Liu et al., 2016), Deeplap_v3 (He et al., 2022), FcnNet (Wu et al., 2022), RetinaNet (Lin et al., 2017), and Mask RCNN were selected as baseline models. The performance index obtained is shown in Table 6 and Figure 12.

TABLE 5 The performance index comparison of different sizes and aspect ratios.

Network parameter	*AP*	AP^{50}	AP^{75}	AP^{s}	AR^{10}	AR^{s}	T-time (s)	P-time(s)
P-**Mask RCNN-A1A2**	**0.636**	**0.877**	**0.689**	**0.724**	**0.727**	**0.727**	**4523.1**	**0.042**
P-Mask RCNN-A1B2	0.628	0.884	0.717	0.713	0.715	0.715	4528.2	0.041
P-Mask RCNN-A1C2	0.612	0.874	0.673	0.699	0.701	0.702	4473.1	0.043
P-Mask RCNN-A1D2	0.632	0.888	0.702	0.709	0.712	0.712	4436.0	0.04
P-**Mask RCNN-B1A2**	**0.605**	**0.862**	**0.650**	**0.688**	**0.693**	**0.693**	**4441.5**	**0.042**
P-**Mask RCNN-B1B2**	**0.647**	**0.862**	**0.710**	**0.721**	**0.724**	**0.724**	**4466.2**	**0.041**
P-**Mask RCNN-B1C2**	**0.614**	**0.860**	**0.666**	**0.716**	**0.718**	**0.718**	**4630.5**	**0.041**
P-**Mask RCNN-B1D2**	**0.641**	**0.891**	**0.733**	**0.728**	**0.732**	**0.732**	**4454.8**	**0.044**
P-Mask RCNN-C1A2	0.615	0.877	0.648	0.709	0.711	0.711	4438.8	0.042
P-Mask RCNN-C1B2	0.632	0.875	0.708	0.718	0.720	0.720	4448.6	0.041
P-Mask RCNN-C1C2	0.627	0.877	0.699	0.712	0.715	0.715	4432.5	0.042
P-Mask RCNN-C1D2	0.635	0.871	0.711	0.717	0.719	0.719	5220.2	0.041
P-Mask RCNN-D1A2	0.641	0.873	0.736	0.732	0.735	0.735	4430.3	0.041
P-Mask RCNN-D1B2	0.642	0.871	0.735	0.728	0.730	0.730	5127.5	0.041
P-Mask RCNN-D1C2	0.633	0.873	0.697	0.720	0.723	0.723	4593.7	0.042
P-Mask RCNN-D1D2	0.634	0.882	0.706	0.720	0.723	0.723	4568.4	0.041
P-Mask RCNN-E1A2	0.628	0.876	0.699	0.714	0.716	0.716	4735.7	0.042
P-**Mask RCNN-E1B2**	**0.625**	**0.867**	**0.701**	**0.713**	**0.716**	**0.715**	**4556.7**	**0.043**
P-Mask RCNN-E1C2	0.634	0.871	0.735	0.718	0.720	0.720	4481.3	**0.04**
P-**Mask RCNN-E1D2**	**0.610**	**0.861**	**0.670**	**0.693**	**0.696**	**0.696**	**4447.3**	**0.041**

Mask RCNN, mask region-based conventional neural network.

Table 6 and Figure 12 show that compared with the SsdNet, Deeplap_v3, FcnNet, and RetinaNet models, Mask RCNN demonstrated the best performance, with an AP^{50} and AR^{s} of approximately 0.805 and 0.583, respectively. Other indicators were also better in Mask RCNN, which proves that Mask RCNN can best solve the problem of overlapped image segmentation. Compared with the Mask RCNN network, the improved Mask RCNN network achieved a better performance, with an AP^{50} and AR^{s} of 0.891 and 0.732, respectively, which represents an improvement of 8.6% and 14.9%, respectively, although there was also an increase in training time and inference time. In summary, the improved Mask RCNN proposed in this study performed the best of the baseline networks studied and can effectively and accurately carry out image segmentation of overlapped tobacco shreds.

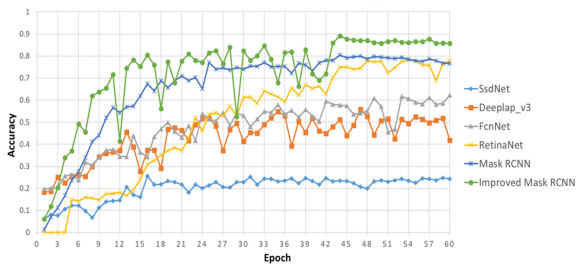


FIGURE 12
Multi-network performance comparison about AP^{50}.

4.3 Evaluation of the COT algorithm

This paper proposes an algorithm for the overlapped region and calculating the overlapped area to obtain an accurate estimation of the overlapped area in the covered tobacco shred, hence overcoming the issues of obtaining the overlapped region in the covered tobacco shred and the failure to detect the overlapped area when different types of tobacco shred overlap. Twenty samples from four tobacco shred types were selected to develop the original tobacco shred sample set, as shown in Table 7. It can be seen that Y-1 to Y-5 are five tobacco silk samples with different shapes, G-1 to G-5 are five different shapes of cut stem samples, P-1 to P-5 are five samples of expanded tobacco silk with different shapes, and Z-1 to Z-5 are five samples of reconstituted tobacco shred with varying shapes. The actual tobacco shred area in each of the 20 samples obtained by the OpenCV algorithm is shown in Table 8.

In this study, 24 various overlapped samples based on the 20 original tobacco shred sample sets were constructed, as shown in Table 9. The abscissa is the serial number (1–4) of the tobacco shred sample, and the vertical coordinate is the type of overlapped tobacco shred (GY, GZ, PG, PY, PZ, and YZ).

The outline of the overlapped region in the covered tobacco shred was obtained, and its area was calculated using the COT algorithm for 24 different random tobacco shred overlap types. The experimental results are shown in Table 10.

Table 10 indicates that among the total area of overlapped tobacco shreds calculated by the OpenCV algorithm, the PG-4 overlapped type sample showed the worst area detection effect, with an AAR of 0.648, whereas the GY-2 overlapped type sample showed the best area detection effect, with an AAR of 0.933. The average actual area ratio detected of overlapped tobacco shred areas was 0.812. However, it can be seen that there is still a large discrepancy between the observed overlapped total area in the tobacco shreds, and the actual area, because there is a missing overlapped area in the covered tobacco shred. The overlapped area obtained by the COT algorithm effectively makes up for the lack of the total area. Moreover, the GY-1 overlapped type had the worst area detection effect, with a CAR of 0.789. The best area detection effect, with a detection rate of 1%, was shown for the GY-3 and GZ-3 overlapped types. The average area detection rate of the overlapped tobacco shred reached 0.90. The worst and best area detection increase rates of CAR compared with AAR were 1.8% and 15.8%, respectively. The average area detection increase rate was 8.8%. In addition, the COT algorithm was applied to 24 experimental sets, and no negative optimization occurred, showing that the algorithm had excellent performance and was effective.

5 Conclusion

This study develops an improved Mask RCNN segmentation model with a COT algorithm to overcome the problems of having

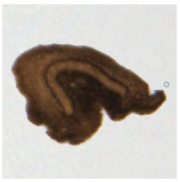
TABLE 6 The performance index comparison of different instance segmentation methods.

Network structure	AP	AP^{50}	AP^{75}	AP^{s}	AR^{10}	AR^{s}	T-time (s)	P-time (s)
SsdNet	0.111	0.246	0.082	0.111	0.337	0.362	1187.2	0.01
Deeplap_v3	0.368	0.558	0.312	0.351	0.491	0.491	2616.4	0.024
FcnNet	0.391	0.623	0.347	0.383	0.532	0.532	2047.0	0.026
RetinaNet	0.416	0.783	0.370	0.416	0.575	0.589	2484.4	0.033
Mask RCNN	0.466	0.805	0.476	0.582	0.583	0.583	3725.4	0.035
Improved Mask RCNN	**0.641**	**0.891**	**0.733**	**0.728**	**0.732**	**0.732**	**4454.8**	**0.044**

Mask RCNN, mask region-based conventional neural network.

10.3389/fpls.2023.1108560

TABLE 7 The original tobacco shred sample datasets.

Type of sample	1	2	3	4	5
Y					
G					
P					
Z					

G, cut stem; P, expanded tobacco silk; Y, tobacco silk; Z, reconstituted tobacco shred.

many types of overlapped tobacco shreds, difficulty in the segmentation of small overlapped tobacco shred objects, and obtaining overlapped region and area calculation. The proposed model can be used to calculate the area of overlapped parts in tobacco shred images. Within the study context, this model was successfully applied to the instance segmentation and area calculation of overlapped tobacco shreds. Based on the aforementioned statements, the following innovations were achieved:

1. A database of 920 overlapped tobacco shred images and two original overlapped tobacco shred image datasets was developed to segment overlapped tobacco shred types, effectively avoiding overfitting and ensuring suitability for actual field use.
2. An improved Mask RCNN network was proposed by adopting DenseNet121 as the backbone, adding upsampling and connection with C2 and C3 level as U-FPN structure, and optimizing anchors parameters, effectively improving the segmentation accuracy for overlapped tobacco shred images. The DenseNet121 model improves the ability of the Mask RCNN network to extract tiny features in the shallow information of overlapped tobacco shreds. The utilization rate of the shallow information and extracted tiny features of tobacco shreds was enhanced with the U-FPN structure. anchors parameters were optimized to reduce both the failure to detect tobacco shreds and redundant calculations.

TABLE 8 The actual area of tobacco shred in each of the 20 samples.

Area (category)	1	2	3	4	5
Y	6611	8952	9747.5	8243	5956
G	7650	9692.5	7359.5	9813	11227
P	3917.5	7227.5	4999	4861	6277.5
Z	7305	5289.5	8593.5	5464.5	6706.5

G, cut stem; P, expanded tobacco silk; Y, tobacco silk; Z, reconstituted tobacco shred.

10.3389/fpls.2023.1108560

TABLE 9 The 24 overlapped tobacco shred samples with different overlapped types.

Type of overlapped sample	1	2	3	4
GY				
GZ				
PG				
PY				
PZ				
YZ				

GP, cut stem and expanded tobacco silk; GY, cut stem and tobacco silk; GZ, cut stem and reconstituted tobacco shred; PY, expanded tobacco silk and tobacco silk; PZ, expanded tobacco silk and reconstituted tobacco shred; YZ, tobacco silk and reconstituted tobacco shred.

3. A COT algorithm was proposed to obtain the overlapped tobacco shred region and calculate the overlapped area, which avoids the loss in the overlapped region area. The COT algorithm significantly improves the detection accuracy of the total area of the overlapped tobacco shreds without negative optimization.

Accordingly, the method proposed in this paper can accurately perform image segmentation of overlapped tobacco shreds and area calculation of the overlapped region. However, this study has several limitations, namely that the number of samples within the datasets was insufficient, the segmentation accuracy must be further improved, and the stacked tobacco shred in abnormal tobacco shred was not studied.

TABLE 10 Pixel area of the 24 overlapped tobacco shred samples under different random overlapping conditions.

Serial number	Sample	Tobacco shred 1	Tobacco shred 2	Area_1 + Area_2	AAR	CAR
1	**GY-1**	**G-1**	**Y-2**	**16602.0**	**0.771**	**0.789**
2	**GY-2**	**G-2**	**Y-4**	**17935.5**	**0.933**	**1.05**
3	**GY-3**	**G-4**	**Y-1**	**16424**	**0.922**	**1**
4	GY-4	G-5	Y-2	20179	0.842	0.898
5	GZ-1	G-4	Z-5	16519.5	0.846	0.952
6	GZ-2	G-2	Z-5	16399	0.8	0.884
7	**GZ-3**	**G-1**	**Z-4**	**13114.5**	**0.842**	**1**
8	GZ-4	G-2	Z-1	16997.5	0.795	0.901
9	PG-1	P-4	G-1	12511	0.758	0.857
10	PG-2	P-3	G-4	14812	0.844	0.956
11	PG-3	P-5	G-3	13637	0.844	0.935
12	**PG-4**	**P-2**	**G-2**	**16920**	**0.684**	**0.81**
13	PY-1	P-4	Y-4	13104	0.849	0.916
14	PY-2	P-3	Y-1	11610	0.748	0.799
15	PY-3	P-5	Y-3	16025	0.825	0.947
16	PY-4	P-2	Y-2	16179.5	0.757	0.804
17	PZ-1	P-4	Z-3	13454.5	0.886	0.983
18	PZ-2	P-3	Z-1	12304	0.772	0.852
19	PZ-3	P-5	Z-5	12984	0.719	0.814
20	PZ-4	P-2	Z-2	12517	0.715	0.806
21	YZ-1	Y-2	Z-3	17545.5	0.742	0.802
22	YZ-2	Y-4	Z-5	14949.5	0.871	0.964
23	YZ-3	Y-1	Z-5	13317.5	0.866	0.955
24	YZ-4	Y-2	Z-3	17545.5	0.867	0.917
Average					**0.812**	**0.9**

AAR, actual area ratio; CAR, COT area ratio. G, cut stem; P, expanded tobacco silk; Y, tobacco silk; Z, reconstituted tobacco shred; GG, cut stem and cut stem; GP, cut stem and expanded tobacco silk; GY, cut stem and tobacco silk; GZ, cut stem and reconstituted tobacco shredPY, expanded tobacco silk and tobacco silk; PZ, expanded tobacco silk and reconstituted tobacco shred; YZ, tobacco silk and reconstituted tobacco shred.

Follow-up work should consider the aspects below:

(1) More abnormal tobacco shreds in the production line should be collected to expand the overlapped tobacco shred datasets under different overlapped types.

(2) The effects of different geometric features of tobacco shreds, such as their length, width, area, and aspect ratios, on the segmentation of overlapped tobacco shreds must be explored. These features and image information must be input into the segmentation network to enhance the network performance of the model.

(3) Although the content of the stacked tobacco shreds is small, it still significantly impacts the accuracy of the component determination. In future studies, the stacked tobacco shreds in the abnormally shredded tobacco should also be sent to the datasets for segmentation.

(4) The scheme proposed in this paper must be installed and applied in real-world scenarios to verify the performance of the model and algorithm.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LW and KJ researched the research background, proposed the methodology, designed the experiment, and wrote and revised the manuscript. YF and XX provided experimental materials and equipment, participated in the dataset shooting and designed the

model test experiment. LF, WZ, and YF participated in the overlapping tobacco type experiment and supervised the model verification. QW and QN provided experimental sites and participated in the overall scheme proposal and model modification. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Innovative Funds Plan of Henan University of Technology (No. 2022ZKCJ03); Science and Technology Project of China Tobacco Henan Industrial Co. Ltd. (No. AW202120) and Henan Science and Technology Research Program (No. 201300210100).

Acknowledgments

We thank China National Tobacco Corporation Xuchang Cigarette Factory and Zhengzhou Tobacco Research Institute for providing us with samples of tobacco shred and jointly determining the shooting plan of the samples.

Conflict of interest

The authors declare that this study received funding from Science and Technology Project of China Tobacco Henan Industrial Co. Ltd. (No. AW202120). The funder had the following involvement in the study: design, data collection and analysis. Authors YF, XX, and LF were employed by China Tobacco Henan Industry Co, Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acuña, A. M. (2017). The framework convention on tobacco control of the world health organization. *Rev. Chil. Enferm. Respir.* 33, 180–182. doi: 10.4067/s0717-73482017000300180

Bai-yi, S., Jiong, M., Hao-yang, Y., Hong-jie, W., and jie, Y. (2020). "Detection of ears of rice in field based on SSD," In *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*. 228–232.

Chen, Y., Bin, J., Zou, C., and Ding, M. (2021). Discrimination of fresh tobacco leaves with different maturity levels by near-infrared (NIR) spectroscopy and deep learning. *J. Analytical Methods Chem.* 2021. doi: 10.1155/2021/9912589

Chen, S., Hu, S., Cui, N., Wang, F., Liu, Q., Yu, H., et al. (2015). Effects of' three kinds of wires' blending ratio on cigarette physical and chemical indicators. *Southwest China J. Agric. Sci.* 28 (6), 2742–2745. doi: 10.16213/j.cnki.scjas.2015.06.072

Dong, H., Liu, Y., Liu, F., Zhang, L., Zhou, M. Z., Xia, Y., et al. (2015). *Method for measuring proportion of reconstituted tobacco in cut tobacco on basis of computer vision* (Beijing: China National Intellectual Property Administration). CN104256882A.

Dong, H., Liu, F., Wang, S., Zhou, M., Zhang, L., Liu, Y., et al. (2016a). *Method for measuring ratio of cut stem to cut tobacco based on computer vision* (Beijing: China National Intellectual Property Administration). CN201410448180.

Dong, H., Zhang, L., Zhou, M. Z., Liu, F., Wang, J. P., Xia, Y., et al. (2016b). *Computer vision-based method for measuring proportion of expanded cut leaves in cut tobacco* (Beijing: China National Intellectual Property Administration). CN201410448216.

Fan, M., Zheng, H., Zheng, S., You, C., Gu, Y., Gao, X., et al. (2020). Mass detection and segmentation in digital breast tomosynthesis using 3D-mask region-based convolutional neural network: A comparative analysis. *Front. Mol. Biosci.* 7, 599333. doi: 10.3389/fmolb.2020.599333

Fernández-Campos, M., Huang, Y. T., Jahanshahi, M. R., Wang, T., Jin, J., Telenko, D. E., et al. (2021). Wheat spike blast image classification using deep convolutional neural networks. *Front. Plant Sci.* 12, 1054. doi: 10.3389/fpls.2021.673505

Gao, Z. Y., Wang, A., Dong, H., Liu, Y., Wang, J. P., and Zhou, M. Z. (2017). Identification of tobacco components in cut filler based on convolutional neural network. *Tob. Sci. Technol.* 50, 68–75. doi: 10.16135/j.issn1002-0861.2016.0502

Gong, B., Ergu, D., Cai, Y., and Ma, B. (2020). Real-time detection for wheat head applying deep neural network. *Sensors* 21 (1), 191. doi: 10.3390/s21010191

Hashemi-Beni, L., Gebrehiwot, A., Karimoddini, A., Shahbazi, A., and Dorbu, F. (2022). Deep convolutional neural networks for weeds and crops discrimination from UAS imagery. *Front. Remote Sens.* 3. doi: 10.3389/frsen

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," In *Proceedings of the IEEE international conference on computer vision*. 2961–2969.

He, J., Lyu, D., He, L., Zhang, Y., Xu, X., Yi, H., et al. (2022). Combining object-oriented and deep learning methods to estimate photosynthetic and non-photosynthetic vegetation cover in the desert from unmanned aerial vehicle images with consideration of shadows. *Remote Sens.* 15 (1), 105. doi: 10.3390/rs15010105

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

He, Y., Zhou, Z., Tian, L., Liu, Y., and Luo, X. (2020). Brown rice planthopper (Nilaparvata lugens stal) detection based on deep learning. *Precis. Agric.* 21 (6), 1385–1402. doi: 10.1007/s11119-020-09726-2

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint* arXiv:1704.04861.

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

Huang, L., Wu, X., Peng, Q., and Yu, X. (2021). Depth semantic segmentation of tobacco planting areas from unmanned aerial vehicle remote sensing images in plateau mountains. *J. Spectrosc.* 2021, 1–14. doi: 10.1155/2021/6687799

Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., and Zheng, Y. (2020). Detection and segmentation of overlapped fruits based on optimized mask r-CNN application in apple harvesting robot. *Comput. Electron. Agric.* 172, 105380. doi: 10.1016/j.compag.2020.105380

Kou, X., Zhang, Y., Zhang, H., Zhang, C., and Ding, N. (2021). Detection of blending proportion of cut tobacco stem based on RGB image features. *Food Machinery* 09), 78–82. doi: 10.13652/j.issn.1003-5788.2021.09.012

Li, R., Liu, Y., Li, W., Li, T., and Wang, J. (2019). Study on rapid determination of tobacco blending uniformity by near infrared spectroscopy. *Food Machinery* 05), 83–87. doi: 10.13652/j.issn.1003-5788.2019.05.015

Li, G., Zhen, H., Hao, T., Jiao, F., Wang, D., and Ni, K. (2021). "Tobacco leaf and tobacco stem identification location detection method based on YOLOV3 network," In *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. (IEEE), 29–31.

Lin, J., Chen, Y., Pan, R., Cao, T., Cai, J., Yu, D., et al. (2022). CAMFFNet: A novel convolutional neural network model for tobacco disease image recognition. *Comput. Electron. Agric.* 202, 107390. doi: 10.1016/j.compag.2022.107390

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," In *Proceedings of the IEEE international conference on computer vision*. 2980–2988.

Lin, H., Zhang, X., Cao, F. J., and Zhou, J. (2020). *A method for determining the mixing ratio of expanded tobacco in finished tobacco* (Beijing: China National Intellectual Property Administration). CN110779832A.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., et al. (2016). "Ssd: Single shot multibox detector," in *European Conference on computer vision* (Cham: Springer), 21–37.

Liu, G., Hou, Z., Liu, H., Liu, J., Zhao, W., and Li, K. (2022). TomatoDet: Anchor-free detector for tomato detection. *Front. Plant Sci.* 13, 942875. doi: 10.3389/fpls.2022.942875

Mei, J., Li, Z., Li, J., Li, J., Su, Z., Xu, B., et al. (2021). Components discrimination for formula tobacco based on hyperspectral imaging. *J. Instrumental Anal.* 08), 1151–1157. doi: 10.19969/j.fxcsxb.20110702

Niu, Q., Liu, J., Jin, Y., Chen, X., Zhu, W., and Yuan, Q. (2022). Tobacco shred varieties classification using multi-Scale-X-ResNet network and machine vision. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.962664

Prasetyo, E., Suciati, N., and Fatichah, C. (2020). "A comparison of YOLO and mask r-CNN for segmenting head and tail of fish," in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*. (IEEE), 1–6.

Qi, J., Zhang, J., Meng, Q., Ju, J., and Jiang, H. (2022). Detection of auxiliary equipment in engine room based on improved SSD. *J. Physics: Conf. Ser.* 2173 (1), 012060. doi: 10.1088/1742-6596/2173/1/012060

Shao, H., Tang, R., Lei, Y., Mu, J., Guan, Y., and Xiang, Y. (2021). Rice ear counting based on image segmentation and establishment of a dataset. *Plants* 10 (8), 1625. doi: 10.3390/plants10081625

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556.

State Tobacco Monopoly Administration (2009). *YC/Z 317-2009*. (China: State Tobacco Monopoly Administration).

Su, Z., Wang, Y., Xu, Q., Gao, R., and Kong, Q. (2022). LodgeNet: Improved rice lodging recognition using semantic segmentation of UAV high-resolution remote sensing images. *Comput. Electron. Agric.* 196, 106873. doi: 10.1016/j.compag.2022.106873

Su, H., Wei, S., Wang, M., Zhou, L., Shi, J., and Zhang, X. (2019). "Ship detection based on RetinaNet-plus for high-resolution SAR imagery," in *2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*. (IEEE), 1–5.

Tan, C., Zhang, P., Zhang, Y., Zhou, X., Wang, Z., Du, Y., et al. (2020). Rapid recognition of field-grown wheat spikes based on a superpixel segmentation algorithm using digital images. *Front. Plant Sci.* 11, 259. doi: 10.3389/fpls.2020.00259

Tong, K., Wu, Y., and Zhou, F. (2020). Recent advances in small object detection based on deep learning: A review. *Image Vision Computing* 97, 103910. doi: 10.1016/j.imavis.2020.103910

Wang, D., and He, D. (2022). Fusion of mask RCNN and attention mechanism for instance segmentation of apples under complex background. *Comput. Electron. Agric.* 196, 106864. doi: 10.1016/j.compag.2022.106864

Wei, J., Qi, L., Ma, F., Ding, M, Le, Y, Wang, Y, et al. (2022). Cut tobacco structure detection and cut tobacco component analysis based on image recognition. *J. Light Industry* 37 (3), 82–87.

Wen, D., Chen, W., Yin, J., Song, Y., Ren, M., and Wang, D. (2022). Overlapping bubble detection and tracking method based on convolutional neural network and kalman filter. *Chem. Eng. Sci.* 263, 118059. doi: 10.1016/j.ces.2022.118059

Wen, C., Wu, J., Chen, H., Su, H., Chen, X., Li, Z., et al. (2022). Wheat spike detection and counting in the field based on SpikeRetinaNet. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.821717

Wu, H., Pang, K. K. Y., Pang, G. K. H., and Au-Yeung, R. K. H. (2022). A soft-computing based approach to overlapped cells analysis in histopathology images with genetic algorithm. *Appl. Soft Computing* 126, 109279. doi: 10.1016/j.asoc.2022.109279

Xu, Z., Sun, K., and Mao, J. (2020). "Research on ResNet101 network chemical reagent label image classification based on transfer learning," In *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*. (IEEE), 354–358.

Yang, G., Chen, G., Li, C., Fu, J., Guo, Y., and Liang, H. (2021). Convolutional rebalancing network for the classification of large imbalanced rice pest and disease datasets in the field. *Front. Plant Sci* 12, 671134. doi: 10.3389/fpls.2021.671134

Ye, H., Wang, T., Ding, N., and He, J. (2013). Effects of tracking and mixture on stability of tobacco index. *J. Anhui Agri* 20), 8700–8702. doi: 10.13989/j.cnki.0517-6611.2013.20.028

Yu, Y., Zhang, K., Yang, L., and Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on mask-RCNN. *Comput. Electron. Agric.* 163, 104846. doi: 10.1016/j.compag.2019.06.001

Zhang, Z., Yin, X., and Yan, Z. (2022). Rapid data annotation for sand-like granular instance segmentation using mask-RCNN. *Automation Construction* 133, 103994. doi: 10.1016/j.autcon.2021.103994

Zhang, Y., Zheng, F., Zhang, X., Wang, X., Zhou, S., Zhu, D., et al. (2019). Determination of cut tobacco blending uniformity based on thermogravimetric analysis. *Tobacco Sci. Technol.* 07), 83–88. doi: 10.16135/j.issn1002-0861.2018.0027

Zhong, Y., Zhou, M. Z., Xu, Y., Liu, D. X., Wang, H. Q., Dong, H., et al. (2021). A method for identifying types of tobacco strands based on residual neural network.Tob. *Sci. Technol.* 54, 82–89. doi: 10.16135/j.issn1002-0861.2020.0602

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 May 2023
DOI 10.3389/fpls.2023.1147034

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
Agricultural Research Service, United States Department of Agriculture (USDA), United States

REVIEWED BY
Brandon Weihs,
United States Department of Agriculture (USDA), United States
Yanbo Huang,
United States Department of Agriculture (USDA), United States

*CORRESPONDENCE
Pan Gao
gp_inf@shzu.edu.cn
Xin Lv
lxshz@126.com

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 18 January 2023
ACCEPTED 27 March 2023
PUBLISHED 10 May 2023

CITATION
Huang Y, Yan J, Zhang Y, Ye W, Zhang C, Gao P and Lv X (2023) Automatic segmentation of cotton roots in high-resolution minirhizotron images based on improved OCRNet.
Front. Plant Sci. 14:1147034.
doi: 10.3389/fpls.2023.1147034

COPYRIGHT
© 2023 Huang, Yan, Zhang, Ye, Zhang, Gao and Lv. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automatic segmentation of cotton roots in high-resolution minirhizotron images based on improved OCRNet

Yuxian Huang [1], Jingkun Yan [2], Yuan Zhang [2], Weixin Ye [2], Chu Zhang [3], Pan Gao [2*] and Xin Lv [1*]

[1]College of Agriculture, Shihezi University, Shihezi, China, [2]College of Information Science and Technology, Shihezi University, Shihezi, China, [3]School of Information Engineering, Huzhou University, Huzhou, China

Root phenotypic parameters are the important basis for studying the growth state of plants, and root researchers obtain root phenotypic parameters mainly by analyzing root images. With the development of image processing technology, automatic analysis of root phenotypic parameters has become possible. And the automatic segmentation of roots in images is the basis for the automatic analysis of root phenotypic parameters. We collected high-resolution images of cotton roots in a real soil environment using minirhizotrons. The background noise of the minirhizotron images is extremely complex and affects the accuracy of the automatic segmentation of the roots. In order to reduce the influence of the background noise, we improved OCRNet by adding a Global Attention Mechanism (GAM) module to OCRNet to enhance the focus of the model on the root targets. The improved OCRNet model in this paper achieved automatic segmentation of roots in the soil and performed well in the root segmentation of the high-resolution minirhizotron images, achieving an accuracy of 0.9866, a recall of 0.9419, a precision of 0.8887, an F1 score of 0.9146 and an Intersection over Union (IoU) of 0.8426. The method provided a new approach to automatic and accurate root segmentation of high-resolution minirhizotron images.

KEYWORDS

plant root, image processing, computer vision, semantic segmentation, attention mechanism

1 Introduction

The root system is the nutrient organ of plants, which plays an important role in promoting plant growth. In root studies, the root phenotypic parameter is an important measure of root growth status. Since soil is an opaque medium, it is not possible to survey roots directly to obtain root phenotypic parameters, except by some methods. Excavation is a traditional method of obtaining roots that can expose the roots to the soil for the purpose

of direct root survey and can be divided into methods such as shovelomics method (Trachsel et al., 2011), core method (Wasson et al., 2016) and trench profile method (Faye et al., 2019). But excavation is a destructive sampling method and also requires considerable time and labor. Hydroponics can be selected to cultivate plants in order to monitor their growth status and obtain plant root information (Li et al., 2018). Although root information is more readily available through hydroponic cultivation of plants than through excavation, the findings from studies using hydroponics cannot be generalized to root studies in the soil environments. In order to obtain root information quickly and easily in the soil environments without damaging plants, root researchers have used sensors to monitor plant roots non-destructively, such as X-ray computed tomography (CT) to reconstruct the three-dimensional structure of rice roots (Teramoto et al., 2020), nuclear magnetic resonance (NMR) imaging technology to analyze the root structure of wheat (Pflugfelder et al., 2022), and minirhizotrons to photograph cotton roots (Shen et al., 2020). Although all of these methods can obtain *in situ* non-destructive images of roots, these methods are applicable to different scenarios. The implementation of CT and NMR imaging technology requires expensive and highly technical equipment, and both methods are in most cases only applicable to the observation of plant roots in small potted plants. In contrast, minirhizotrons are less expensive and simpler to operate, and minirhizotrons can be inserted into the soil in the field to observe the roots. Therefore, minirhizotrons are well suited to be used to collect *in situ* images of plant roots in the field to analyze root phenotypic parameters.

Traditional root segmentation is achieved manually by image processing software which is time-consuming and labor-intensive. To address the drawbacks of manual root segmentation, machine learning methods are applied by root researchers for automatic root segmentation. Machine learning is a method that allows machines to simulate or learn human behavior, and common machine learning methods include OTSU (Otsu, 1979), Support Vector Machine (SVM) (Cortes and Vapnik, 1995) and Random Forests (Breiman, 2001). A segmentation method based on color features of the roots of wheat seedlings was implemented: firstly, the root image was converted from RGB color space to HCI color space, then the threshold of the chroma component was set to extract the binary image, and finally, the image was processed with local fuzzy c-means clustering algorithm to get the segmentation result (Goclawski et al., 2009). The OTSU method was applied to the study of automatic root segmentation of the images acquired by desktop scanners, and this method is an image segmentation algorithm based on a dynamic threshold (Chen and Zhou, 2010). The threshold segmentation methods are usually only applicable to the automatic segmentation of root images with simple backgrounds. Moreover, the threshold segmentation methods need to be set with suitable thresholds in advance, which leads to the poor generalization of the threshold segmentation methods.

The Convolutional Neural Network (CNN) is a method proposed to compensate for the previous machine learning's inability to learn autonomously like the human brain. LeNet-5 (Lecun et al., 1998), as one of the earliest CNNs, successfully realized the recognition of handwritten fonts. Later, the CNN-based AlexNet (Krizhevsky et al., 2012) was proposed to achieve the automatic classification of images. The proposal of FCN (Long et al., 2015) gave CNNs the ability of semantic segmentation. Then, excellent semantic segmentation methods such as U-Net (Ronneberge et al., 2015), PSPNet (Zhao et al., 2017), DeepLabv3+ (Chen et al., 2018) and OCRNet (Yuan et al., 2020) were created. With the development of image segmentation algorithms for deep learning, many root researchers have applied CNN-based image segmentation models in automatic root segmentation research. An improved DeepLabv3+ model was used for the automatic segmentation of cotton roots, where root images were acquired from minirhizotrons installed in the field, and experimental results showed that this improved semantic segmentation model worked well for root segmentation of minirhizotron images with a real soil environment as the background (Shen et al., 2020). U-Net was applied to the segmentation of soybean seedling roots, and experimental results showed that the method could achieve accurate segmentation of soybean seedling roots (Xu et al., 2022). CNN-based segmentation methods have been widely used in automatic root segmentation research and have achieved good performance. These methods not only do not need to be set with suitable thresholds in advance but also can be applied to root segmentation in real soil environments. To further improve the performance of semantic segmentation models, many root researchers have added attention mechanisms to the segmentation models to enhance the root segmentation capability of the models.

The attention mechanism is an information processing mechanism that focuses on useful information and ignores useless information. The classical attention mechanisms contain Squeeze-and-Excitation (SE) (Hu et al., 2018), Convolutional Block Attention Module (CBAM) (Woo et al., 2018) and Non-Local (Wang et al., 2018). The recently proposed attention mechanisms contain Efficient Channel Attention (ECA) (Wang et al., 2020), Coordinate Attention (CA) (Hou et al., 2021) and Global Attention Mechanism (GAM) (Liu et al., 2021). Since the attention mechanism can enhance the focus of the segmentation model on the root targets, the attention mechanism has been applied in some root segmentation studies for improving the segmentation model and enhancing the root segmentation capability of the model. For example, an improved U-Net with the SE attention module was applied to the study of rice root segmentation, which achieved automatic precision segmentation of rice seedling roots in the images (Gong et al., 2021).

Nowadays, most of the root segmentation researchers still use DeepLabv3+ and the previous semantic segmentation models. As a model proposed in recent years, OCRNet uses the object region to which each pixel belongs as the region for extracting contextual representation, which is a better way to obtain contextual representation for each pixel than DeepLabv3+. Moreover, the background of *in situ* images is a real soil environment, and there are many noises in the soil that interfere with automatic root segmentation, and these noises will affect the accuracy of automatic root segmentation. An attention mechanism can be added to the model, which can enhance the model's attention to

the root targets and improve the model's ability to distinguish the roots from the background. Therefore, this work aims to explore the application of an advanced semantic segmentation network model improved by an attention mechanism for cotton root segmentation of minirhizotron images with a real soil environment as the background. The specific objectives achieved here are as follows:

(1) Collect high-resolution *in situ* cotton root images using minirhizotrons and annotate these images.
(2) Improve OCRNet by adding the GAM attention module to optimize the pixel representations output by the backbone.
(3) Compare and evaluate the improved method with the mainstream semantic segmentation methods.

2 Materials and methods

2.1 Data collection

The high-resolution minirhizotron images were collected at the experimental station of Shihezi University College of Agriculture in Shihezi, Xinjiang Uygur Autonomous Region (85°59′43.7064″E, 44° 19′21.1044″N). In 2018, we selected two areas in the cotton field, installed two minirhizotrons in each area, and collected data eight times (July 6, July 10, July 14, July 18, July 22, July 26, July 30, and August 3). Each minirhizotron was used to scan cotton roots at three different depths at a time. We acquired a total of 96 high-resolution minirhizotron images, each with a size of 2271 × 2550 pixels.

2.2 Data annotation

We screened the high-quality images among these 96 high-resolution minirhizotron images. In this process, we removed four images that were blurred, leaving 92 high-quality root images. We used the LabelMe 3.16.7 (Russell et al., 2008) annotation tool to annotate these 92 high-resolution minirhizotron images and generate corresponding annotated images, with the roots marked in red and the background in black (Figure 1). The average annotation time per image was 8 hours.

2.3 Data augmentation

In order to expand the dataset and improve the generalization ability of the model obtained from subsequent training, we performed data augmentation on these 92 high-resolution minirhizotron images after data annotation was completed. We used five ways of data augmentation, namely, increasing luminance, decreasing luminance, isometric enlargement, isometric reduction, and adding salt-and-pepper noise. Finally, the number of images was expanded to 552, and the number of annotated images was also expanded to 552 accordingly to form the final high-resolution minirhizotron image dataset. We generated 460 new annotated images from 92 annotated images by corresponding transformation operations according to the data augmentation, so the 460 new images generated from 92 original images by data augmentation do not need to be manually annotated again. We divided the 552 high-resolution minirhizotron images into the training set, the validation set and the test set in the ratio of 6: 2: 2. There are 330 images in the training set, 111 images in the validation set and 111 images in the test set. The training set was used for training the network model, and the validation set was used to select the model weights that performed best during model training to evaluate the model performance by using the test set.

2.4 Segmentation model

OCRNet is a semantic segmentation method that makes the classification of each pixel and the segmentation of each class more accurate by augmenting the representation of each pixel with the object-contextual representation (OCR) (Yuan et al., 2020). We applied the OCRNet semantic segmentation model to the root segmentation of high-resolution minirhizotron images. In the semantic segmentation model, the backbone plays the role of



FIGURE 1
Data Annotation. **(A)** Original image. **(B)** Annotated image.

representation extraction which has an important impact on the performance of the segmentation model.

HRNetV2 (Sun et al., 2019a) is a network that retains high-resolution representations well and is well suited for segmentation of high-resolution images with elongated or tiny objects. In the high-resolution minirhizotron image dataset we acquired, the resolution of each image is relatively high, and the roots in each image are very fine, so the model used must be able to notice the minute details of the roots to achieve accurate segmentation of it. Therefore, HRNetV2 is well suited for extracting root representations from high-resolution minirhizotron images. We used HRNetV2 as the backbone of OCRNet to extract pixel representations. HRNetV2 contains four stages with four parallel subnetworks. In HRNetV2, the resolution is gradually decreased to a half and accordingly the number of channels is increased to the double. The first stage contains 4 residual units where each unit, the same to the ResNet-50 (He et al., 2016), is formed by a bottleneck with 64 channels, and is followed by one 3×3 convolution reducing the number of channels of feature maps to C (C represents the number of channels of the high-resolution subnetworks in the last three stages). The 2nd, 3rd and 4th stages contain 1, 4 and 3 exchange blocks, respectively. One exchange block contains 4 residual units where each unit contains two 3 × 3 convolutions in each resolution and an exchange unit across resolutions. In summary, there are a total of 8 exchange units, i.e., 8 multi-scale fusions are conducted. Unlike HRNetV1 (Sun et al., 2019b), HRNetV2 concatenates the representations of the four resolution subnetworks in the last stage, making full use of the representations of each resolution subnetwork. The network structure of HRNetV2 is shown in Figure 2. We chose HRNetV2-W48 as the final backbone, where 48 represents the number of channels (C) of the high-resolution subnetworks in the last three stages. The number of channels in the other three parallel subnetworks is 96, 192 and 384 for HRNetV2-W48.

After extracting the pixel representations by using HRNetV2-W48 as the backbone of OCRNet, a coarse semantic segmentation result is output by FCN to obtain K soft object regions (soft object region refers to the region consisting of pixels of each class in the coarse semantic segmentation result). K equals 2 because there are only two classes in the annotation of our high-resolution minirhizotron images, i.e., background and root. After obtaining the soft object regions, the pixel representations output from the backbone and the semantic representations of the corresponding K soft object regions are weighted and summed to obtain the K object region representations. The calculation of each object region representation is shown as follows:

$$f_k = \sum_{i \in I} \tilde{m}_{ki} x_i \tag{1}$$

where f_k is the representation of the kth object region. x_i is the representation of pixel i. I refers to the set of pixels in the image. $\tilde{m}_{ki}$ is the normalized degree for pixel i belonging to the kth object region.

Then, the pixel-region relation is obtained by computing the relation between each pixel and each object region according to the self-attention (Vaswani et al., 2017), which is shown as follows:

$$w_{ik} = \frac{e^{k(x_i,\, f_k)}}{\sum_{j=1}^{K} e^{k(x_i,\, f_j)}} \tag{2}$$

where w_{ik} is the relation between the ith pixel and the kth object region. $k(x_i, f_i) = \varnothing(x)^{\top}\psi(f)$ is the unnormalized relation function, $\varnothing(\cdot)$ and $\psi(\cdot)$ are two transformation functions consisting of a 1×1 convolution layer, a Batch Normalization (BN) layer and a ReLU function. K is the number of object regions.

After obtaining the pixel-region relation, the object contextual representations are calculated based on the pixel-region relation and the object region representations. The calculation of the object contextual representation for each pixel is shown as follows:

$$y_i = \rho\left(\sum_{k=1}^{K} w_{ik}\delta(f_k)\right) \tag{3}$$

where y_i is the object contextual representation of pixel i. $\rho(\cdot)$ and $\delta(\cdot)$ are two transformation functions consisting of a 1×1 convolution layer, a BN layer and a ReLU function. K is the number of object regions. w_{ik} is the relation between the ith pixel and the kth object region. f_k is the representation of the kth object region.

Finally, the augmented representation for each pixel is calculated as the aggregation of the original representation for each pixel and the object contextual representation for each pixel, which is shown as follows:

$$z_i = g([x_i^{\top} y_i^{\top}]^{\top}) \tag{4}$$

where z_i is the augmented representation for pixel i. $g(\cdot)$ is a transformation function used to fuse the original representation and the object contextual representation, consisting of a 1×1

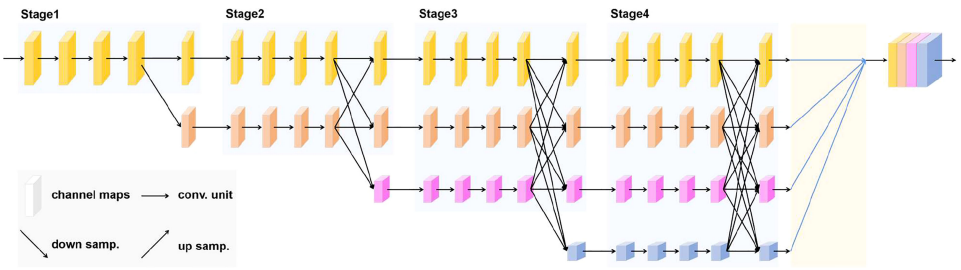


FIGURE 2
HRNetV2 network structure.

convolution layer, a BN layer and a ReLU function. x_i is the representation of pixel i . y_i is the object contextual representation of pixel i. The whole OCR module takes the pixel representations output by the backbone as input and the augmented representations as output, as illustrated in Figure 3. After obtaining the augmented representations, the output with the number of classes equal to the number of channels is obtained by a 1×1 convolution layer, and then the output is restored to the original scale by bilinear up-sampling to obtain the final semantic segmentation prediction result.

2.5 Model improvement

GAM is a global attention mechanism that focuses on the interaction of the three dimensions of widths, heights, and the number of channels in the feature maps (Liu et al., 2021). Therefore, GAM reduces information reduction and magnifies global dimension-interactive features, which allows the network model to focus on the features of the targets in a comprehensive manner. In our collected high-resolution minirhizotron images, the color of the roots is similar to the color of the soil in the background, which causes the model to have more difficulty in distinguishing the roots from the soil in the background. And the images we collected include a variety of complex background noises, such as stones, worms, soil cracks, residual plastic film, etc. These background noises can interfere with the model in distinguishing the roots from the background. And enhancing the model's focus on the roots is a key way to improve the model's ability to distinguish the roots from the background. Because of the role of making the model to focus on the features of the targets in a comprehensive manner, the addition of the GAM module in the model can enhance the model's focus on the roots to improve the model's ability to distinguish the roots from the background. Therefore, we improved OCRNet by adding the GAM module. And by this improvement, the model's ability to identify and segment the roots can be improved. GAM follows the structure of CBAM in which the channel attention submodule and spatial attention submodule are connected in series. But GAM redesigned the channel attention submodule and the spatial attention submodule in the structure. In GAM, given an input feature map, GAM outputs a three-dimensional channel attention map through the channel attention submodule, multiplies the input and output of the channel attention submodule to obtain the input of the spatial attention submodule, and then outputs a three-dimensional spatial attention map, and multiplies the input and output of the spatial attention submodule to obtain the final output. The overall attention process can be summarized as:

$$F_2 = M_C(F_1) \otimes F_1 \tag{5}$$

$$F_3 = M_S(F_2) \otimes F_2 \tag{6}$$

where $\otimes$ denotes element-wise multiplication. F_1 is the feature map input by GAM and $F_1 \in \mathbb{R}^{C\times H\times W}$. M_C is the output of the channel attention submodule of GAM and $M_C \in \mathbb{R}^{C\times H\times W}$. F_2 is the result of multiplying M_C and F_1. M_S is the output of the spatial attention submodule of GAM and $M_S \in \mathbb{R}^{C\times H\times W}$. F_3 is the final output of GAM.

In the channel attention submodule of GAM, the 3D information of the feature map is retained by a 3D permutation module, magnified by a two-layer Multi-Layer Perceptron (MLP), converted into 3D information in the original dimensional order by a 3D reverse permutation module, and finally input into a sigmoid function to obtain the channel attention map. The structure of the channel attention submodule of GAM is shown in Figure 4.

In the spatial attention submodule of GAM, the number of channels of the feature map is first reduced to C/r by a 7×7 convolutional layer, followed by a 7×7 convolutional layer to restore the number of channels to C, and finally input into a sigmoid function to obtain the spatial attention map. The structure of the spatial attention submodule of GAM is shown in Figure 5.

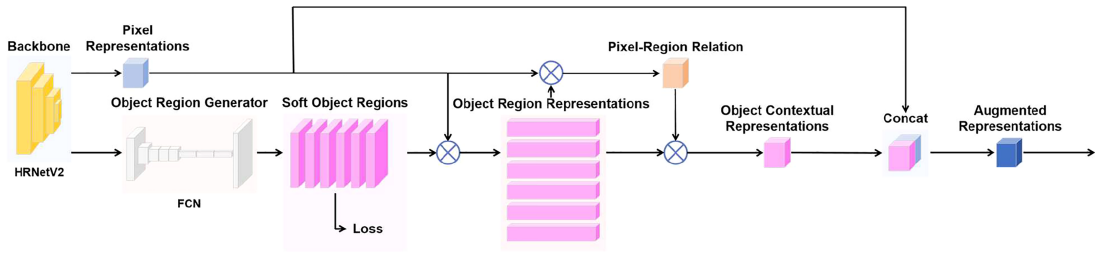


FIGURE 3
OCR module.

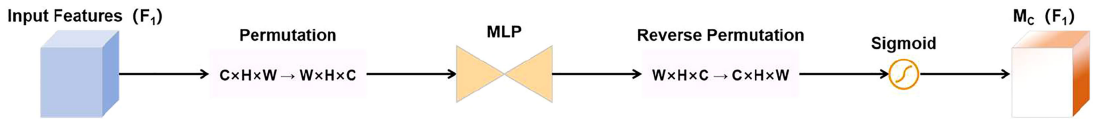


FIGURE 4
Channel attention submodule of GAM.

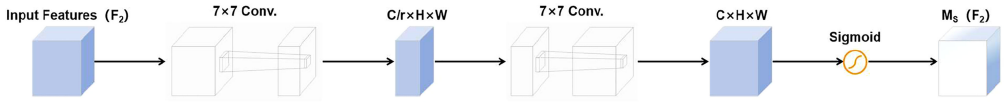


FIGURE 5
Spatial attention submodule of GAM.

The pixel representations are involved in the computation several times throughout the OCR module, affecting the individual outputs in the module (Figure 5). It shows that the pixel representations are closely related to the segmentation effect of OCRNet. In order to optimize the pixel representations and thus improve the segmentation effect of OCRNet, we improved the OCR module by adding the GAM module. In the improved OCR module, the pixel representations are input to the GAM module, and the GAM module outputs the optimized pixel representations. Instead of the original pixel representations, the optimized pixel representations participate in the computation of the object region representations, the pixel-region relation, and the object contextual representations. The improved OCR module is shown in Figure 6.

2.6 Network training

The OCRNet we designed has two outputs, one is the coarse semantic segmentation result output by using FCN as the object region generator, and the other is the final prediction result output by the whole model. For both outputs, we used two pixel-wise cross-entropy loss functions to calculate these two loss values separately, with the loss weight set to 0.4 for the former and 1 for the latter. Since the root segmentation is a pixel-level binary classification problem, the calculation of the cross-entropy loss is shown as follows:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \ln p_i + (1-y_i)\ln(1-p_i)] \quad (7)$$

where N is the number of pixels, y_i is the label of pixel i, and p_i is the predicted probability value.

Due to the limited GPU memory and the fact that the model parameters, gradients, optimizer states and intermedia activations cost the GPU memory during training, the size of the images input to the network model during training cannot be too large. During the training process, we set up a random crop pipeline so that the high-resolution original images were randomly cropped into sub-images of size 512 × 512 pixels before being input into the network model. We used the polynomial decay method to achieve the decay of the learning rate, the initial learning rate was set to 0.01, the power of the polynomial was set to 0.9, and the minimum learning rate was set to 0.0001. In order to make the model training more stable and converge better, we used the SGDM optimizer with momentum set to 0.9 and weight decay set to 0.0005. We set the batch size to 4 and the total number of iterations to 40,000. In the process of training the network model, we saved the model weights every 500 iterations, tested the model weights using the validation set, and selected the best-performing model weights to evaluate the model performance by using the test set. The detailed hyperparameters during network training are shown in Table 1.

The server environment was Windows 10, and the program was compiled and run in Python 3.7. The model was trained, validated, and tested under PyTorch 1.8.1 and CUDA 11.1. The server was

TABLE 1 Hyperparameters during network training.

Hyperparameters	Value
Loss function	Cross-entropy loss
Random crop size	512 × 512 pixels
Input size	512 × 512 pixels
Learning rate decay strategy	Polynomial decay
Initial learning rate	0.01
Power of polynomial decay	0.9
Minimum learning rate	0.0001
Optimizer	SGDM
Momentum of SGDM	0.9
Weight decay	0.0005
Batch size	4
Max iterations	40000

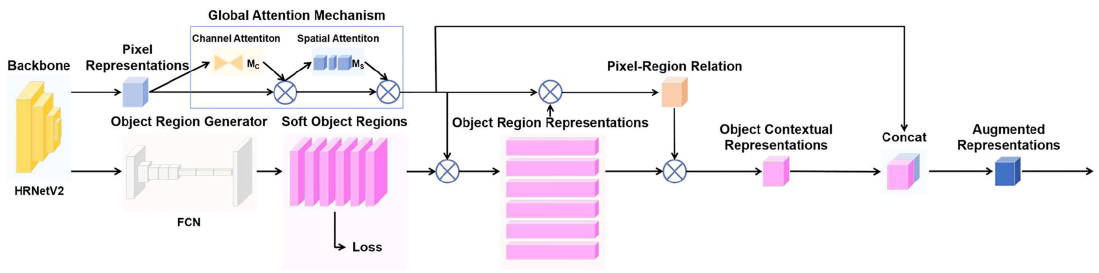


FIGURE 6
Improved OCR module.

equipped with an NVIDIA GeForce RTX 3080 Laptop (16G) graphics card for model training acceleration.

2.7 Evaluation

In order to objectively and reasonably evaluate the root segmentation performance of our model, we utilized five evaluation metrics, i.e., accuracy, recall, precision, F1 score, and Intersection over Union (IoU):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{11}$$

$$\text{IoU} = \frac{TP}{TP+FP+FN} \tag{12}$$

In Equations 8, 9, 10, 11 and 12, the *TP*, *FP*, *FN*, and *TN* denote the true positive (the area which is both predicted and annotated as root area), false positive (the area which is predicted as root area but annotated as background), false negative (the area which is predicted as background but annotated as root area) and true negative (the area which is both predicted and annotated as background) measurements. Accuracy is the proportion of the number of correctly predicted samples to the total number of samples. Recall is the proportion of samples that are correctly predicted as positive cases to all samples with true labels as positive cases. Precision is the proportion of samples correctly predicted as positive cases to all samples predicted as positive cases. F1 score is the harmonic mean of precision and recall. IoU is a commonly used measure in semantic segmentation to evaluate the overlap ratio of predicted results to ground truth.

3 Results and analysis

3.1 Training set loss and validation set loss

We set the number of model training iterations to 40,000 and used our high-resolution minirhizotron image dataset to train our improved OCRNet model. The total training time of the network model is 23.5 h. The training set loss is shown in Figure 7A. With the increasing number of training iterations, the training set loss decreased in the general trend. After 35,000 iterations, the training set loss was stabilized within 0.03, which indicates that the training set loss converged and the model was well trained. The validation set loss is shown in Figure 7B. With the increasing number of training iterations, the validation set loss also decreased under the general trend. Although the fluctuation of the validation set loss was a bit large in the middle two periods, it was within the reasonable range of the fluctuation of the validation set loss. After 35000 iterations, the validation set loss was stabilized within 0.029, which indicates that the validation set loss converged and stabilized.

3.2 Segmentation performance of improved OCRNet

In the process of training the network model, we saved the model weights every 500 iterations and tested the model weights using the validation set. Since IoU is one of the most common evaluation metrics in semantic segmentation, we selected the model weights with the highest IoU values for root segmentation in three randomly selected high-resolution minirhizotron images and compared the segmented images with the manually annotated images, as shown in Figure 8. Our improved OCRNet had good segmentation performance for root segmentation of the high-resolution minirhizotron images. In the segmentation results of these three randomly selected high-resolution minirhizotron images, many taproots and lateral roots were accurately identified and segmented. The overall segmentation results are very close to the annotated images.

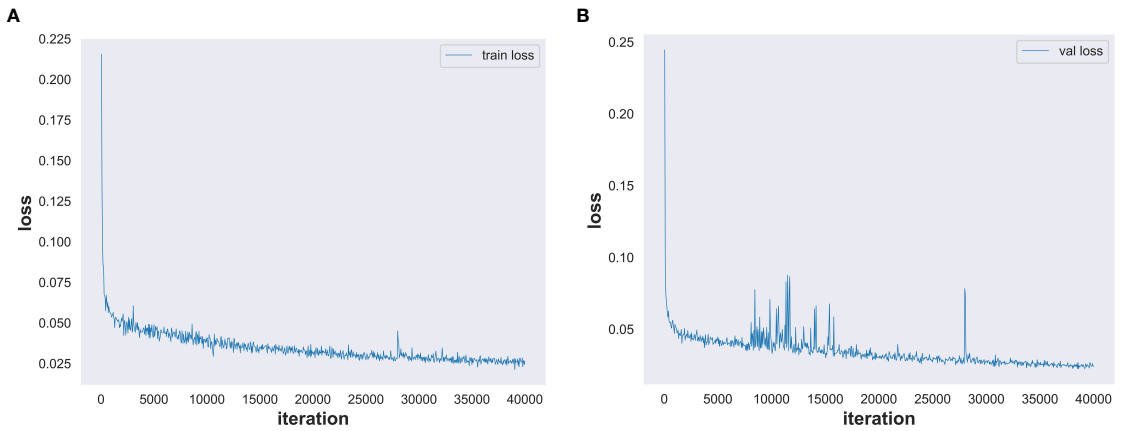


FIGURE 7
Training set loss and validation set loss. **(A)** Training set loss. **(B)** Validation set loss.

10.3389/fpls.2023.1147034

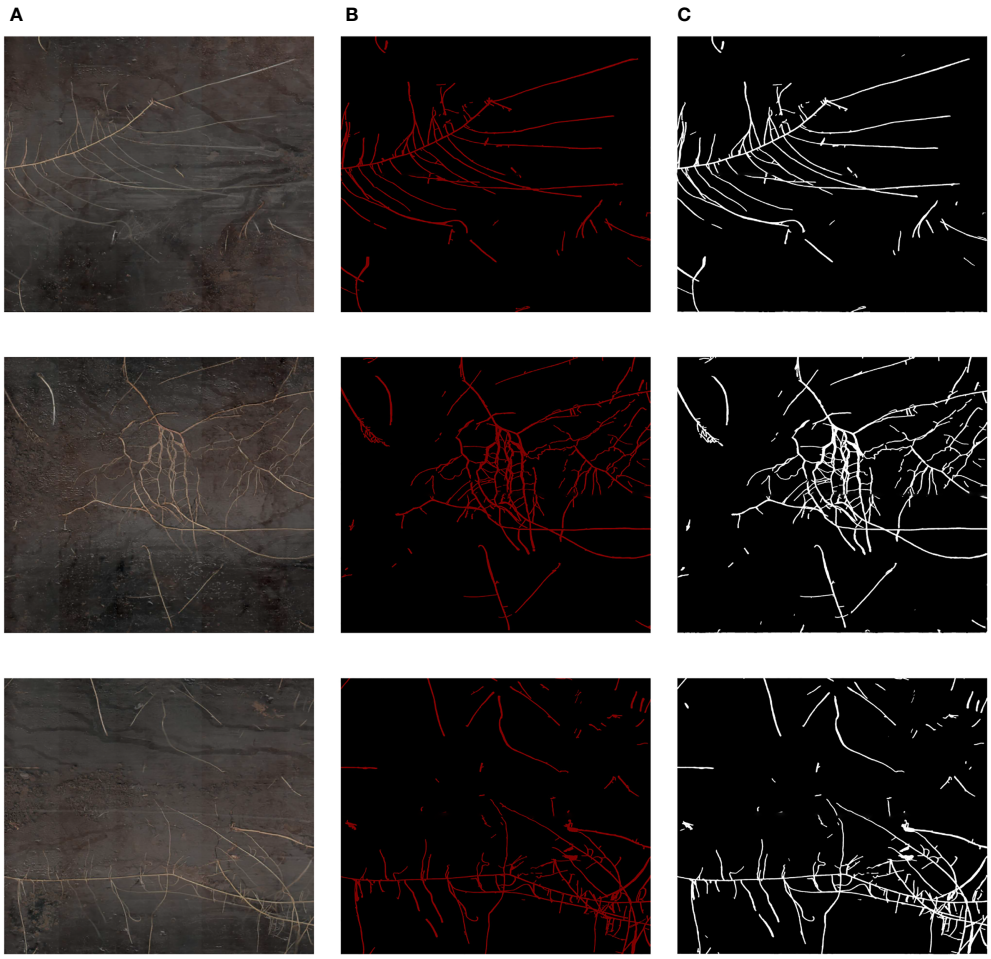


FIGURE 8
Comparison of our improved OCRNet segmented images and the corresponding annotated images. **(A)** Original images. **(B)** Annotated images. **(C)** Improved OCRNet segmented images.

3.3 Comparison with mainstream segmentation methods

To compare our improved method with the current mainstream semantic segmentation methods, we selected the original OCRNet, DeepLabv3+, PSPNet and FCN as comparison models, and the hyperparameters of each model during training were the same as those set during the training of our improved OCRNet. We used HRNetV2-W48 as the backbone of the original OCRNet and ResNet-50-C (He et al., 2019) as the backbone of DeepLabv3+, PSPNet and FCN. All comparison models were trained by using the high-resolution minirhizotron image dataset we produced. Similarly, we set the model weights to be saved and tested using the validation set every 500 iterations. The model weights with the highest IoU values for each semantic segmentation method when being tested using the validation set were used to evaluate the model segmentation performance by using the test set. As shown in Table 2, the five best model weights were used to obtain the values of each evaluation metric for each segmentation method, and the values of each evaluation metric were accurate to four decimal places. Although the precision values of our improved OCRNet model are slightly lower than that of DeepLabv3+, our improved OCRNet has the highest accuracy values, recall values, F1 score values and IoU values. The precision is the proportion of samples correctly predicted as positive cases to all samples predicted as positive cases. If there is a more serious under-segmentation of the model's root segmentation results, then the precision may be high instead. Therefore, the precision cannot determine how well a model performs root segmentation. The IoU is a commonly used measure in semantic segmentation to evaluate the overlap ratio of predicted results to ground truth, so it is better than precision to evaluate the segmentation performance of a model. And our improved OCRNet achieved the highest IoU value of 0.8426. It illustrates that our improved semantic segmentation method has a stronger segmentation capability.

As shown in Figure 9, we selected a high-resolution minirhizotron image with very intricate roots to show the results of root segmentation of this image by using the five semantic segmentation methods in Table 2. As a whole, the integrity of the roots segmented in this image by each method is relatively high. Due to the high resolution of this root image, in Figure 9, we marked a red box in the original image as well as in the same

TABLE 2 Comparison between the different methods.

Method	Backbone	Accuracy	Recall	Precision	F1 score	IoU
FCN	ResNet-50-C	0.9826	0.9344	0.8511	0.8908	0.8031
PSPNet	ResNet-50-C	0.9826	0.9367	0.8495	0.8901	0.8034
DeepLabv3+	ResNet-50-C	0.9852	0.9121	**0.8949**	0.9034	0.8239
OCRNet	HRNetV2-W48	0.9859	0.9330	0.8875	0.9097	0.8343
Improved OCRNet	HRNetV2-W48	**0.9866**	**0.9419**	0.8887	**0.9146**	**0.8426**

Bold values indicates the maximum values in their columns.

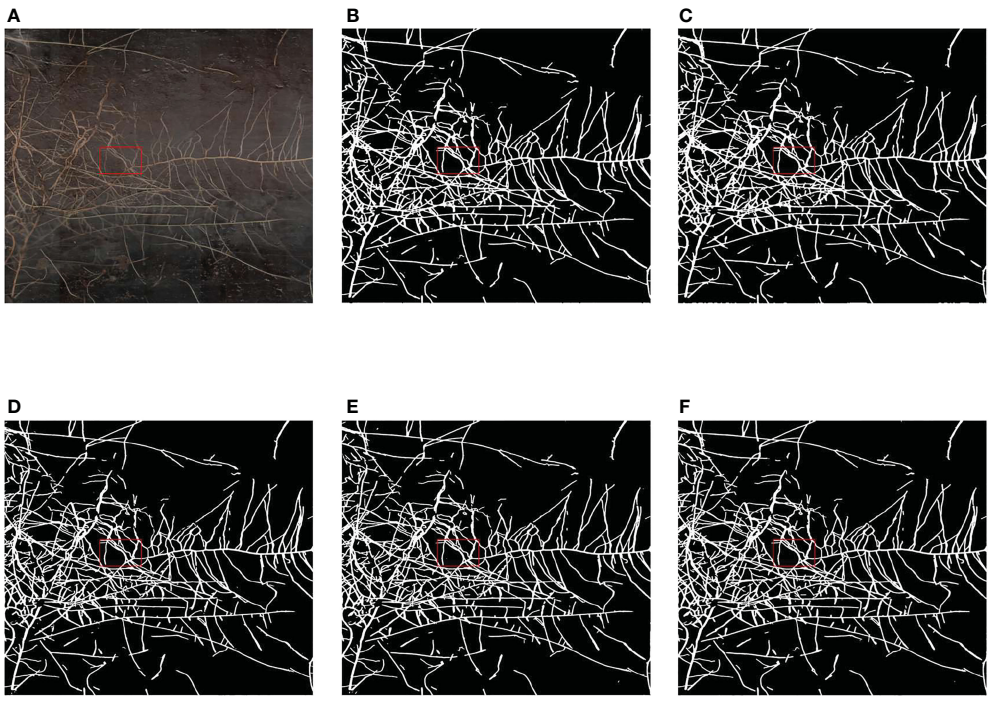


FIGURE 9
Comparison of the segmentation results of the five methods. **(A)** Original image. **(B)** Segmentation result of FCN. **(C)** Segmentation result of PSPNet. **(D)** Segmentation result of DeepLabv3+. **(E)** Segmentation result of OCRNet. **(F)** Segmentation result of improved OCRNet.

location of each segmented image. Then, we zoomed in on the marked red box area of each segmented image to compare the segmentation details of each method, as shown in Figure 10.

In Figure 10, we marked six regions in the original image. At the position corresponding to region 1 in the original image, the root segmentation results of FCN and PSPNet were over-segmented, and the root contour segmented by DeepLabv3+ was very rough. At the position corresponding to region 2 in the original image, the root segmentation results of FCN and PSPNet were very rough because FCN and PSPNet identified some background parts of the two root gaps as root parts. At the position corresponding to region 3 in the original image, the root segmentation result of DeepLabv3+ was under-segmented. At the position corresponding to region 4 in the original image, the root segmentation results of FCN, PSPNet and DeepLabv3+ were under-segmented, and the root segmentation result of OCRNet was slightly over-segmented. At the position corresponding to region 5 in the original image, OCRNet identified the background noise as the root. At the position corresponding to region 6 in the original image, none of the five semantic segmentation methods segmented the gap between the roots very accurately, among which FCN and PSPNet did not identify the gap part between the roots basically. From the segmentation results of the six regions, there were more under-segmented regions in the root segmentation results for DeepLabv3+. This suggests that there may also be many under-segmented regions in the root segmentation results of DeepLabv3+ for other images, which may be the reason why the precision values of DeepLabv3+ are higher in Table 2. Overall, our improved OCRNet can pay attention to more root details and has better segmentation performance.

3.4 Segmentation performance in complex backgrounds

The root images we collected include a variety of complex background noises, such as stones, worms, soil cracks, residual

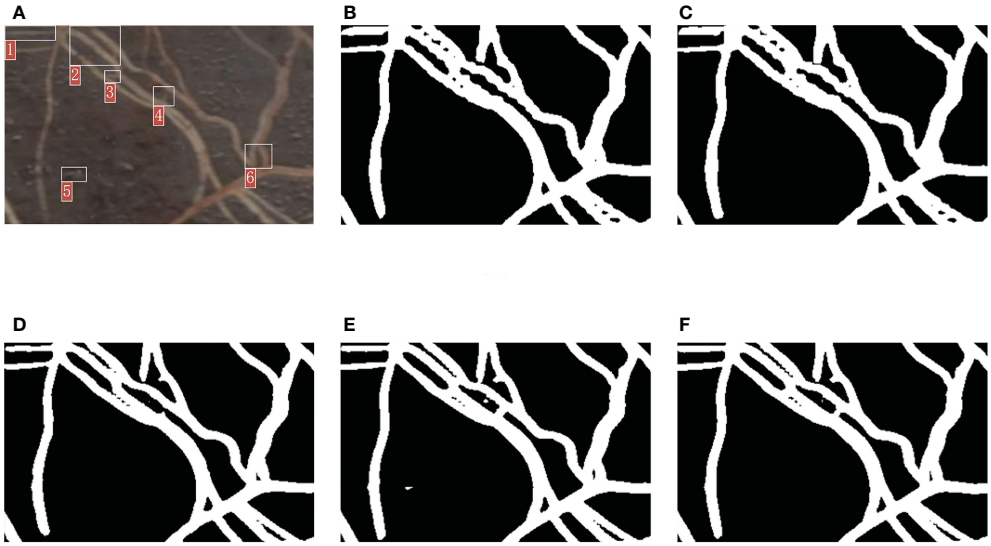


FIGURE 10
Comparison of the segmentation details of the five methods. **(A)** Original image. **(B)** Segmentation result of FCN. **(C)** Segmentation result of PSPNet. **(D)** Segmentation result of DeepLabv3+. **(E)** Segmentation result of OCRNet. **(F)** Segmentation result of improved OCRNet.

plastic film, etc. We selected some regions of the root images with these background noises to show the results of our method to segment these regions, as shown in Figure 11. In Figure 11A, this region contains background noise such as stones, and in Figure 11B, our method segmented the roots of this region accurately and was largely unaffected by the stones in the background. In Figure 11C, this region contains a white worm, and in Figure 11D, although our method was also able to segment the roots in this region, it identified the body of this white worm as a root as well, which may be caused by the similarity of the body size and color of such white worms to the roots. In Figure 11E, there are some soil cracks in this region, and the shape of both the cracks and the roots are bar-shaped, and in Figure 11F, our method accurately separated the roots from the soil cracks. In Figure 11G, the left side of this region contains the residual plastic film in the soil, and in Figure 11H, our method classified only a very small portion of the residual plastic

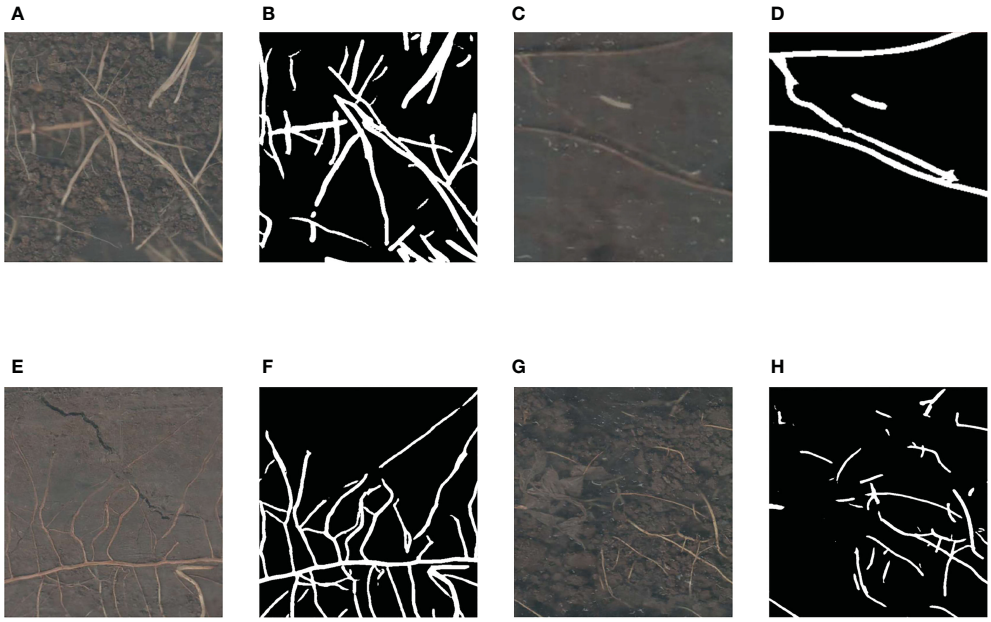


FIGURE 11
Various complex noises in the soil. **(A)** The region with stones in the background. **(B)** Segmentation result of the region with stones in the background. **(C)** The region with a white worm in the background. **(D)** Segmentation result of the region with a white worm in the background. **(E)** The region with some soil cracks in the background. **(F)** Segmentation result of the region with some soil cracks in the background. **(G)** The region with residual plastic film in the background. **(H)** Segmentation result of the region with residual plastic film in the background.

film as roots. Therefore, our method was able to distinguish the roots from other types of complex background noise relatively accurately, except for not identifying white worms as background noise. We believe that the small number of samples containing white worms in the background is also the reason why our trained model was unable to learn enough such cases to distinguish between roots and white worms correctly.

4 Discussion

The root images acquired in the real environments contain a variety of complex background noises. And these complex background noises can seriously interfere with automatic root identification and segmentation. Therefore, many root segmentation studies usually adopt the method of cultivating roots in ideal laboratory environments to exclude these interfering factors (Krzyzaniak et al., 2021; Xu et al., 2022; Zhao et al., 2022). The root images taken in ideal laboratory environments have less complex background noise, which is more favorable for root image segmentation. But the segmentation models trained by using the root images taken in ideal laboratory environments cannot meet the requirements for root segmentation in real soil environments. Minirhizotrons can be used to obtain *in-situ* root images non-destructively (Xu et al., 2020). To make our research applicable to actual production environments, we collected root images by minirhizotrons in a real soil environment.

The original images collected by minirhizotrons are basically high-resolution (Shen et al., 2020; Bauer et al., 2022). And there is usually not enough memory in the GPU to load high-resolution images for training due to the limited GPU memory (Gong et al., 2021). If the high-resolution images are resized so that the high-resolution images can be input into the network model for training, the resolution of the images will be reduced, and many details of the roots will be lost. It will prevent the model from learning enough root representations, resulting in a decrease in the model's ability to distinguish the roots from the background. To solve this problem, we added a random cropping pipeline so that the root images were randomly cropped into sub-images of size 512 × 512 pixels before being input into the model for training. By adding a random cropping pipeline to reduce the size of the input images, the images were able to be input into the model for training while retaining the root details. However, in order to obtain high-level semantic information, in the process of extracting representations, the usual semantic segmentation model will first obtain low-resolution representations through down-sampling, and then the low-resolution representations will be restored to high-resolution representations by up-sampling, and this approach will lose a lot of valid information in the process of up-sampling and down-sampling (Sun et al., 2019b). Therefore, we used HRNetV2 as the backbone of the segmentation model to extract the root representations, which retained the high-resolution root representations through the parallel multi-resolution subnetworks in HRNetV2.

The traditional method of manually segmenting the roots in the images is very inefficient. Not only the speed of manual root segmentation is slow, but also the results of manual root segmentation may not be completely correct due to the visual fatigue problem in manual root segmentation (Abramoff et al., 2004; Le Bot et al., 2010). Therefore, an accurate and rapid root segmentation method is needed to replace inefficient manual root segmentation (Smith et al., 2020). In this study, our improved OCRNet model achieved automatic segmentation of roots in the soil and performed well in the root segmentation of the high-resolution minirhizotron images we acquired, achieving an accuracy of 0.9866, a recall of 0.9419, a precision of 0.8887, an F1 score of 0.9146 and an IoU of 0.8426, as shown in Table 2. And our improved OCRNet has the highest accuracy values, recall values, F1 score values and IoU values. It indicates the superiority of our improved method. It can be seen from Figure 8 that the root segmentation results of our improved method are very close to the manually annotated images. Moreover, our method takes about 0.3 seconds to segment a root image of size 2271 × 2550 pixels, while we manually annotate the same root image in about 8 hours. It shows that our method is much faster than manual annotation. In terms of the accuracy of segmentation and the time taken for segmentation, our method has basically reached the level of replacing manual annotation.

Although our method has been able to achieve automatic root segmentation of high-resolution minirhizotron images taken under a real soil environment, it has some shortcomings in root segmentation. After comparing and analyzing the original images and the segmented images, we found that our method did not segment the filamentous roots and the light-colored roots accurately enough. As shown in Figure 12, we selected a region containing the filamentous roots and a region containing the light-colored roots to show the segmentation results of our method for these regions. In Figure 12A, this region contains many filamentous roots which are very fine and occupy very few pixels. In Figure 12B, the root parts in the segmentation result of our method for this part of the region were generally wider than those in the original image, and our method did not identify some very fine roots in the region. In our opinion, because of the small pixel area occupied by the filamentary roots, the model did not learn enough samples of these roots, which led to inaccurate segmentation results for these roots. In Figure 12C, this region contains many light-colored roots which are buried by the soil, so these roots are lightly colored in the image. In Figure 12D, our method did not identify some light-colored roots. It may be that the characteristics of the light-colored roots are not obvious, which makes it difficult for the model to accurately distinguish these roots from the background.

In the future, we will continue to collect and annotate more *in situ* cotton root images, especially those containing complex background noise, to expand our dataset. By expanding the dataset, the model will learn more samples, which will improve the generalizability of the model and the model's ability to distinguish the roots from the background. Meanwhile, we will continue to improve our root segmentation model already improved in this paper to solve the problem of difficult segmentation of the filamentous roots and the light-colored roots by enhancing the superiority of the model framework.

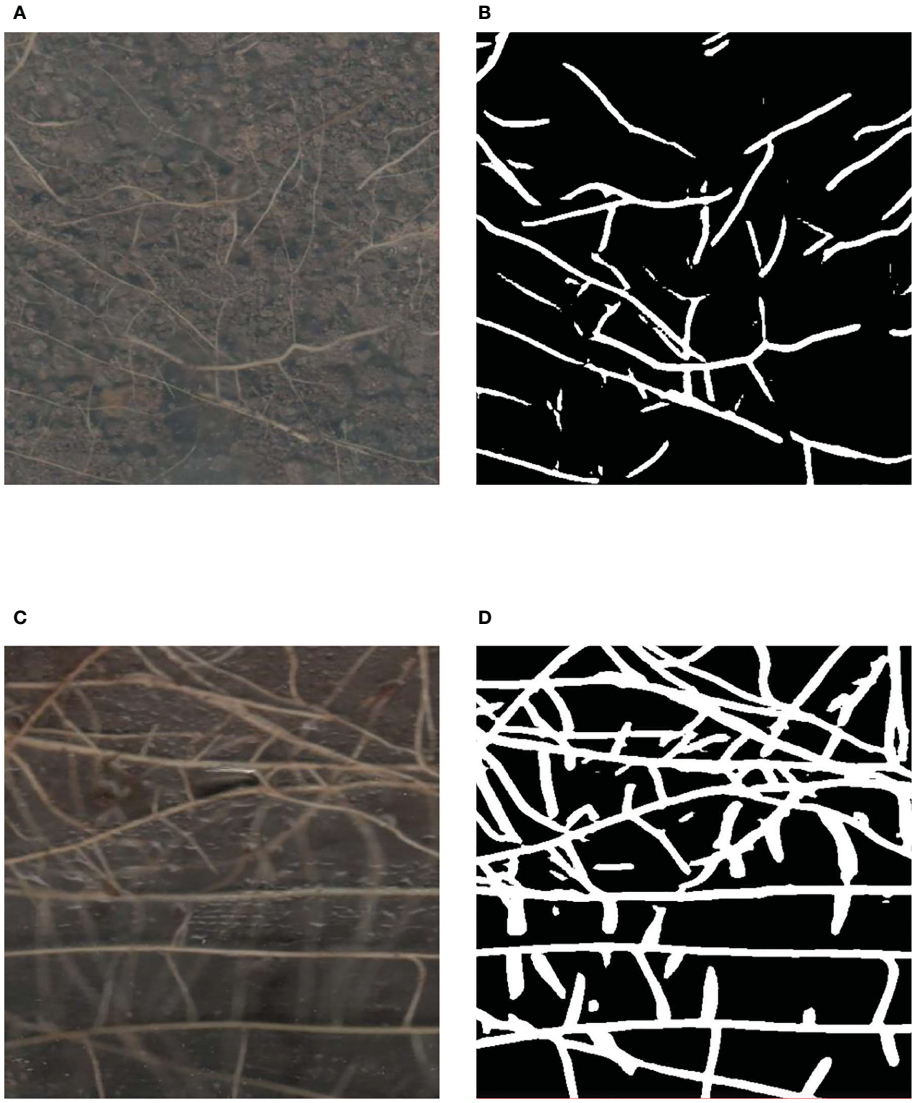


FIGURE 12
The filamentous root region and the light-colored root region. **(A)** The filamentous root region. **(B)** Segmentation result of the filamentous root region. **(C)** The light-colored root region. **(D)** Segmentation result of the light-colored root region.

5 Conclusion

To solve the problem of low efficiency of traditional manual root segmentation and to achieve automatic root segmentation of high-resolution minirhizotron images taken in a real soil environment, we improved the OCRNet by adding the GAM attention module and achieved accurate automatic segmentation of cotton roots using the improved OCRNet. Firstly, in order to make our research applicable to actual production environments, we collected images of roots in a real soil environment using minirhizotrons and produced a dataset. Then, we applied the OCRNet to the root segmentation of high-resolution minirhizotron images and selected HRNetV2 as the backbone of the OCRNet. Meanwhile, the structure of the original OCRNet was improved by adding the GAM attention module after the pixel representations output by HRNetV2, which made the pixel representations augmented. The pixel representations augmented by the attention mechanism were used to participate in the calculation of relevant parameters in the OCR module, which subsequently improved the ability of the model to distinguish the roots from the background. Next, we trained our improved OCRNet by using the high-resolution minirhizotron image dataset and set up a random cropping pipeline to preserve the details in the high-resolution minirhizotron images within the GPU memory limit. Finally, our improved OCRNet model achieved automatic segmentation of roots in the soil and performed well in the root segmentation of the high-resolution minirhizotron images we acquired, achieving an accuracy of 0.9866, a recall of 0.9419, a precision of 0.8887, an F1 score of 0.9146 and an IoU of 0.8426, as shown in Table 2. And our improved OCRNet has the highest accuracy values, recall values, F1 score values and IoU values. This method provided a new approach to automatic and accurate root segmentation of high-resolution minirhizotron images taken in the soil environments, which laid the foundation for automatic analysis of root phenotypic parameters in the field of root research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YH, JY, and YZ conceived the idea. PG, and XL contributed to the preparation of equipment and acquisition of data. YH wrote the code and tested the method. YH, and YZ validated the results. YH wrote the manuscript. JY, WY, CZ, PG, and XL revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (NO.61965014) and the International Cooperation Science and Technology Project of Shihezi University (GJHZ202205).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abramoff, M. D., Magalhaes, P. J., and Ram, S. J. (2004). Image processing with image. *Biophotonics Int.* 11, 36–42. doi: 10.3233/ISU-1991-115-601

Bauer, F. M., Larm, L., Morandage, S., Lobet, G., Vanderborght, J., Vereecken, H., et al. (2022). Development and validation of a deep learning based automated minirhizotron image analysis pipeline. *Plant Phenomics* 2022. doi: 10.34133/2022/9758532

Breiman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi: 10.1023/a:1010933404324

Chen, Y., and Zhou, X. (2010). "Plant root image processing and analysis based on 2D scanner," in *Fifth international conference on bio-inspired computing: theories and applications, BIC-TA 2010, University of Hunan, Liverpool Hope University, Liverpool*, United Kingdom / Changsha, China; 2010 September 8-10 and September 23-26; (Piscataway, NJ: IEEE).

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer vision - ECCV 2018. ECCV 2018. lecture notes in computer science*, vol. 11211 . Eds. V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (Cham: Springer), 833–851. doi: 10.1007/978-3-030-01234-2_49

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20 (3), 273–297. doi: 10.1023/a:1022627411411

Faye, A., Sine, B., Chopart, J. L., Grondin, A., Lucas, M., Diedhiou, A. G., et al. (2019). Development of a model estimating root length density from root impacts on a soil profile in pearl millet (*Pennisetum glaucum* (L.) r. br). application to measure root system response to water stress in field conditions. *PloS One* 14 (7), 1–18. doi: 10.1371/journal.pone.0214182

Goclawski, J., Sekulska-Nalewajko, J., Gajewska, E., and Wielanek, M. (2009). An automatic segmentation method for scanned images of wheat root systems with dark discolourations. *Int. J. Appl. Mathematics Comput. Sci.* 19 (4), 679–689. doi: 10.2478/v10006-009-0055-x

Gong, L., Du, X. F., Zhu, K., Lin, C. H., Lin, K., Wang, T., et al. (2021). Pixel level segmentation of early-stage in-bag rice root for its architecture analysis. *Comput. Electron. Agric.* 186, 106197. doi: 10.1016/j.compag.2021.106197

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, Las Vegas, NV, USA; 2016 June 27-30; (Piscataway, NJ: Computer Vision Foundation / IEEE), 770–778. doi: 10.1109/CVPR.2016.90

He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., and Li, M. (2019). "Bag of tricks for image classification with convolutional neural networks," in *2019 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019,* Long Beach, CA, USA; 2019 June 16-20; (Piscataway, NJ: Computer Vision Foundation / IEEE), 558–567. doi: 10.1109/CVPR.2019.00065

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *2021 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual*; 2021 June 19-25; (Piscataway, NJ: Computer Vision Foundation / IEEE), 13708–13717. doi: 10.1109/CVPR46437.2021.01350

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-Excitation networks," in *2018 IEEE Conference on computer vision and pattern recognition, CVPR 2018*. Salt Lake City, UT, USA; 2018 June 18-22; (Piscataway, NJ: Computer Vision Foundation / IEEE), 2011–2023.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks," in *Advances in neural information processing systems 25: 26th annual conference on neural information processing systems 2012. Proceedings of a meeting held December 3-6, 2012*, Lake Tahoe, Nevada, United States. (New York: Curran Associates, Inc. / Red Hook, NY), 1097–1105.

Krzyzaniak, Y., Cointault, F., Loupiac, C., Bernaud, E., Ott, F., Salon, C., et al. (2021). *In situ* phenotyping of grapevine root system architecture by 2D or 3D imaging: Advantages and limits of three cultivation methods. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.638688

Le Bot, J., Serra, V., Fabre, J., Draye, X., Adamowicz, S., and Pages, L. (2010). DART: a software to analyse root system architecture and development from captured images. *Plant Soil* 326 (1-2), 261–273. doi: 10.1007/s11104-009-0005-2

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86 (11), 2278–2324. doi: 10.1109/5.726791

Li, Q. S., Li, X. Q., Tang, B., and Gu, M. M. (2018). Growth responses and root characteristics of lettuce grown in aeroponics, hydroponics, and substrate culture. *Horticulturae* 4 (4). doi: 10.3390/horticulturae4040035

Liu, Y., Shao, Z., and Hoffmann, N. (2021). Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv*. Available at: 10.48550/arXiv.2112.05561.

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *2015 IEEE conference on computer vision and pattern recognition, CVPR 2015*, Boston, MA, USA; 2015 June 7-12; (Piscataway, NJ: Computer Vision Foundation / IEEE), 3431–3440.

Otsu, N. (1979). Threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernetics* 9 (1), 62–66. doi: 10.1109/tsmc.1979.4310076

Pflugfelder, D., Kochs, J., Koller, R., Jahnke, S., Mohl, C., Pariyar, S., et al. (2022). The root system architecture of wheat establishing in soil is associated with varying elongation rates of seminal roots: quantification using 4D magnetic resonance imaging. *J. Exp. Bot.* 73 (7), 2050–2060. doi: 10.1093/jxb/erab551

Ronneberge, O., Fischer, P., and Brox, T. (2015). "U-Net: convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. lecture notes in computer science*, vol. 9351 . Eds. N. Navab, J. Hornegger, W. Wells and A. Frangi (Cham: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vision* 77 (1-3), 157–173. doi: 10.1007/s11263-007-0090-8

Shen, C., Liu, L. T., Zhu, L. X., Kang, J., Wang, N., and Shao, L. M. (2020). High-throughput *in situ* root image segmentation based on the improved DeepLabv3 +Method. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.576791

Smith, A. G., Petersen, J., Selvan, R., and Rasmussen, C. R. (2020). Segmentation of roots in soil with U-net. *Plant Methods* 16 (1), 13. doi: 10.1186/s13007-020-0563-0

Sun, K., Xiao, B., Liu, D., and Wang, J. (2019b). "Deep high-resolution representation learning for human pose estimation," in *2019 IEEE conference on computer vision and pattern recognition, CVPR 2019*, Long Beach, CA, USA; 2019 June 16-20; (Piscataway, NJ: Computer Vision Foundation / IEEE), 5686–5696. doi: 10.1109/CVPR.2019.00584

Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., et al. (2019a). High-resolution representations for labeling pixels and regions. *arXiv*. Available at: https://arxiv.org/abs/1904.04514.

Teramoto, S., Takayasu, S., Kitomi, Y., Arai-Sanoh, Y., Tanabata, T., and Uga, Y. (2020). High-throughput three-dimensional visualization of root system architecture of rice using X-ray computed tomography. *Plant Methods* 16 (1), 66. doi: 10.1186/s13007-020-00612-6

Trachsel, S., Kaeppler, S. M., Brown, K. M., and Lynch, J. P. (2011). Shovelomics: high throughput phenotyping of maize (Zea mays l.) root architecture in the field. *Plant Soil* 341 (1-2), 75–87. doi: 10.1007/s11104-010-0623-8

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Proceedings of the 30th annual conference on neural information processing systems*. (New York: Red Hook, NY), 5998–6008.

Wang, X., Girshick, R., Gupta, A., and He, K. (2018). "Non-local neural networks." in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, Salt Lake City, UT, USA; 2018 June 18-22; (Piscataway, NJ: Computer Vision Foundation / IEEE), 7794–7803. doi: 10.1109/CVPR.2018.00813

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "ECA-net: Efficient channel attention for deep convolutional neural networks." in *2020 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2020*, Seattle, WA, USA; 2020 June 13-19; (Piscataway, NJ: Computer Vision Foundation / IEEE), 11531–11539. doi: 10.1109/CVPR42600.2020.01155

Wasson, A., Bischof, L., Zwart, A., and Watt, M. (2016). A portable fluorescence spectroscopy imaging system for automated root phenotyping in soil cores in the field. *J. Exp. Bot.* 67 (4), 1033–1043. doi: 10.1093/jxb/erv570

Woo, S., Park, J., Lee, J., and Kweon, I. S. (2018). "CBAM: Convolutional block attention module," in *European Conference on computer vision*, vol. 11211 . Eds. V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (Cham: Springer), 3–19. doi: 10.1007/978-3-030-01234-2_1

Xu, X. Y., Qiu, J. K., Zhang, W., Zhou, Z., and Kang, Y. (2022). Soybean seedling root segmentation using improved U-net network. *Sensors* 22 (22). doi: 10.3390/s22228904

Xu, W. H., Yu, G. H., Zare, A., Zurweller, B., Rowland, D., Reyes-Cabrera, J., et al. (2020). Overcoming small minirhizotron datasets using transfer learning. *Comput. Electron. Agric.* 175, 105466. doi: 10.1016/j.compag.2020.105466

Yuan, Y., Chen, X., and Wang, J. (2020). "Object-contextual representations for semantic segmentation," in *Computer vision – ECCV 2020. ECCV 2020. lecture notes in computer science*, vol. 12351 . Eds. A. Vedaldi, H. Bischof, T. Brox and J.-M. Frahm (Cham: Springer), 173–190. doi: 10.1007/978-3-030-58539-6_11

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in *22017 IEEE conference on computer vision and pattern recognition, CVPR 2017*, Honolulu, HI, USA; 2017 July 21-26; (Piscataway, NJ: Computer Vision Foundation / IEEE), 6230–6239. doi: 10.1109/CVPR.2017.660

Zhao, H. J., Wang, N., Sun, H. C., Zhu, L. X., Zhang, K., Zhang, Y. J., et al. (2022). RhizoPot platform: A high-throughput *in situ* root phenotyping platform with integrated hardware and software. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1004904

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 June 2023
DOI 10.3389/fpls.2023.1165552

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Dimitrios Kateris,
Centre for Research and Technology Hellas (CERTH), Greece
Ronnie Concepcion II,
De La Salle University, Philippines

*CORRESPONDENCE
Dedong Kong
kongdd@zaas.ac.cn

RECEIVED 14 February 2023
ACCEPTED 10 May 2023
PUBLISHED 02 June 2023

CITATION
Ye Z, Tan X, Dai M, Lin Y, Chen X, Nie P, Ruan Y and Kong D (2023) Estimation of rice seedling growth traits with an end-to-end multi-objective deep learning framework.
Front. Plant Sci. 14:1165552.
doi: 10.3389/fpls.2023.1165552

COPYRIGHT
© 2023 Ye, Tan, Dai, Lin, Chen, Nie, Ruan and Kong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Estimation of rice seedling growth traits with an end-to-end multi-objective deep learning framework

Ziran Ye[1], Xiangfeng Tan[1], Mengdi Dai[1], Yue Lin[2], Xuting Chen[1], Pengcheng Nie[3], Yunjie Ruan[3,4] and Dedong Kong[1]*

[1]Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2]Institute of Spatial Information for City Brain (ISICA), Hangzhou City University, Hangzhou, China, [3]Institute of Agricultural Bio-Environmental Engineering, College of Bio-systems Engineering and Food Science, Zhejiang University, Hangzhou, China, [4]Academy of Rural Development, Zhejiang University, Hangzhou, China

In recent years, rice seedling raising factories have gradually been promoted in China. The seedlings bred in the factory need to be selected manually and then transplanted to the field. Growth-related traits such as height and biomass are important indicators for quantifying the growth of rice seedlings. Nowadays, the development of image-based plant phenotyping has received increasing attention, however, there is still room for improvement in plant phenotyping methods to meet the demand for rapid, robust and low-cost extraction of phenotypic measurements from images in environmentally-controlled plant factories. In this study, a method based on convolutional neural networks (CNNs) and digital images was applied to estimate the growth of rice seedlings in a controlled environment. Specifically, an end-to-end framework consisting of hybrid CNNs took color images, scaling factor and image acquisition distance as input and directly predicted the shoot height (SH) and shoot fresh weight (SFW) after image segmentation. The results on the rice seedlings dataset collected by different optical sensors demonstrated that the proposed model outperformed compared random forest (RF) and regression CNN models (RCNN). The model achieved R^2 values of 0.980 and 0.717, and normalized root mean square error (NRMSE) values of 2.64% and 17.23%, respectively. The hybrid CNNs method can learn the relationship between digital images and seedling growth traits, promising to provide a convenient and flexible estimation tool for the non-destructive monitoring of seedling growth in controlled environments.

KEYWORDS

growth traits, fresh weight, rice seedling, deep learning, convolution neural network

1 Introduction

Plant factories achieve stable and efficient growing of plants by controlling the growing environment (Ares et al., 2021; McClements et al., 2021). Various plant factories have been promoted in China to cope with the shortage of cultivated lands for vegetable production. In recent years, industrial rice seedlings have attracted attention because of good economic benefits (Ma et al., 2014). At present, the rice seedlings bred in the factory need to be selected manually and then transplanted into the field. Plant growth is a response to environmental parameters (Chen et al., 2016). Plant phenotype is the character of plants under the interaction between intrinsic genotype and external environmental conditions (Furbank and Tester, 2011). Phenotypic morphological traits such as height, leaf area and biomass, can be obtained by measurement and weighing, which is helpful for quantifying plant growth (Watt et al., 2020). The traditional methods of manual trait measurement are simple and accurate, but they are difficult to meet the demand of high-throughput trait acquisition in large quantities, and usually require destructive sampling, which is time-consuming and laborious (Hüther et al., 2020). The estimation of plant growth is a non-negligible element in the intelligent development of plant factories; thus it is of great practical significance to develop rapid, accurate and automatic methods for obtaining plant growth-related traits to replace some tedious manual measurements.

The development of computer vision provides a good opportunity for image-based automatic measurement and acquisition of plant phenotype data. Mortensen et al. (2018) proposed a method for segmenting lettuce in 3D point clouds and estimating their yield. Reyes-Yanes et al. (2020) used MASK-RCNN to segment the lettuce from the background and used the geometric features extracted from the segmented data to build a fresh weight regression model. Nowadays, RGB images can be obtained at a low cost by using sensors such as digital cameras and smartphones, which are affordable and easy to operate. Computer vision algorithms are then used to extract image-based phenotypic data and apply them to downstream tasks. For example, Yu et al. (2013) proposed a crop segmentation method and used the skeleton endpoint to characterize the leaf of the seedling to recognize the growth stage of the seedling. Borianne et al. (2018) developed software for automatic cereal root system measurements from digital images. These works show that images are promising to provide a non-destructive and convenient access to obtain plant growth information, and the key is to construct appropriate feature extraction methods.

In recent years, convolutional neural networks (CNNs), an advanced deep learning method, have been widely applied to visual tasks in the field of agriculture, such as plant detection (Quan et al., 2019), classification (Perugachi-Diaz et al., 2021), segmentation (Gong et al., 2021) and counting (Osco et al., 2020; He et al., 2022). Benefiting from the ability of automatic feature learning and hierarchical feature extraction, CNNs originally designed for classification tasks can also perform well for regression tasks. Some studies have used regression CNNs for plant growth trait estimation in controlled environments, and most of them target lettuce (Chen et al., 2016; Zhang et al., 2020; Buxbaum et al., 2022; Lin et al., 2022) and Arabidopsis thaliana (Ubbens and Stavness, 2017). Existing studies illustrate that regression CNNs can model the relationship between images and growth traits of leafy vegetables well. Meanwhile, there is a great demand for automatic measurement of growth-related traits in rice seedlings (Lu et al., 2017; Mekawy et al., 2018; Zhang et al., 2019). However, applying existing regression CNNs for the growth monitoring of grain crops such as rice still needs to be further validation, as morphological differences between monocot and dicot plants exist from the seedling stage, which challenges the estimation of rice seedling growth traits directly from digital images.

The objective of this study was to accurately estimate growth-related traits of rice seedlings in controlled environment agriculture. A CNN-based framework including image preprocessing, image augmentation, semantic segmentation network and regression network, was used to segment RGB images of rice seedlings and model the relationship between the images and the corresponding growth-related traits (height and fresh weight). This study explored the potential of using CNNs with digital images to estimate growth-related traits of rice seedlings in vertical planting modules to establish a feasible and robust seedling growth monitoring method.

2 Material and methods

2.1 Image collection and preprocessing

The rice cultivar ZY-18 (Zheyou 18, hybrid indica) was selected for experiments. ZY-18 was bred by the Zhejiang Academy of Agricultural Sciences (ZAAS) and has been widely planted in Zhejiang and surrounding provinces in China. The seeds used in this study were obtained from the market. The rice seedlings used in this work were grown in a vertical growth unit in a laboratory in Binjiang, Hangzhou (N30°11′, E120°12′). After surface disinfection, rice seeds were germinated in dark for two days and then sowed into substrate trays. Rice seedlings were grown under controlled climate conditions, with day/night temperatures of 26-28°C/18-20°C and average relative humidity of 75%. During the seedling growth period, full spectrum led grow lights were used for illumination at a light intensity of 400 μmol·m-2·s-1 and 14 to 16 hours during the day. The experiment was performed from November 20, 2021, to December 16, 2021.

A Nikon Z5 camera and a smartphone (iPhone 12) were used for image acquisition, in which the Nikon camera shot 10 seedlings at a time and the iPhone shot one seedling at a time. During the image collection, the sensors were placed on the top of a photography light box (60×60×60 cm) to capture digital images. According to the difference in the sensor size, the resolution of the original digital image is 4016×6016 (Nikon Z5) and 3024×4032 (iPhone), respectively. All digital images are stored in PNG format. Finally, two datasets were constructed, a digital image dataset containing 92 images captured by a digital camera, and a digital image dataset containing 76 images captured by a smartphone.

For the digital images of seedlings in both datasets, binarized labels consistent with the original image size were generated by manual annotation. The rectangle enclosing a single seedling was generated using binarized labels, and the original images were cropped using these rectangles. A total of 984 rice seedling images were obtained as the new dataset.

Referring to the previous study, the image dataset was randomly divided into a training set and a test set in a ratio of 8:2, which were used to construct and evaluate the model, respectively. Further, 20% of the training dataset was randomly selected as the validation set during training. To prevent overfitting, data augmentation was performed on the fly when training all models, which consisted of horizontal flipping, horizontal shift and random rotation.

2.2 Measurement of traits

Manual measurements were taken at the same time as image collection. Ten seedlings were randomly sampled from each planting tray. After washing and drying the seedlings, the plant height and fresh weight were measured using a ruler and a digital balance with a resolution of 0.1 cm and 0.0001 g, respectively. These measurements were conducted on days 10 (December 3), 17 (December 10) and 22(December 15) after seeding. Finally, a regression dataset with 504 samples was obtained, in which each sample had measurements of seedling height and shoot fresh weight after root removal. Meanwhile, each seedling sample had its corresponding binarized label in the segmentation dataset above.

2.3 Construction of the network architecture

The whole process is shown in Figure 1. In the first stage, the RGB image of rice seedlings with the size of 512×512×3 was input into the U-Net (Ronneberger et al., 2015) segmentation model to output the probability map of rice seedlings. In the second stage, based on the pre-experiment, a modified ResNet50 (He et al., 2016) was used as the backbone of the regression network to predict growth-related traits. Specifically, the RGB image of rice seedlings and segmentation predictions above were concatenated as the input of the feature extraction network. Besides, a branch fully connected layer was introduced to receive a geometric vector including the input scaling factor and image acquisition distance as the input. The fully connected layers of the two paths were fused into a feature vector of 576×1 at the depths of the network. And this feature vector was passed to the regression head, which consisted of two fully connected layers. Finally, the regression network output two values, which represented the prediction results of seedling height and fresh weight.

In the training phase, the Adam optimizer was used to optimize the parameters of the two networks in stages. In the beginning, only the parameters of the segmentation network were updated. After 30 epochs, the segmentation model reached convergence and the parameters were frozen. Subsequently, the parameters of the regression network were updated until the model converged. The initial learning rate was set to 0.001, batch size was set to 4, and the training period was set to 300 epochs. The loss function of segmentation and regression were cross entropy loss and smooth L1 loss, respectively. The "ReduceLROnPlateau" scheduler and "Early Stopping" strategy was adopted to adjust the learning rate and control the training process: If the validation loss did not improve within 50 epochs, the learning rate will decrease by 0.1 times. If the validation loss still did not improve within 100 epochs, the training will be terminated.

2.4 Performance evaluation

To evaluate the proposed model, overall accuracy (OA), F1-score and Intersection-over-Union (IoU) metrics were used as the

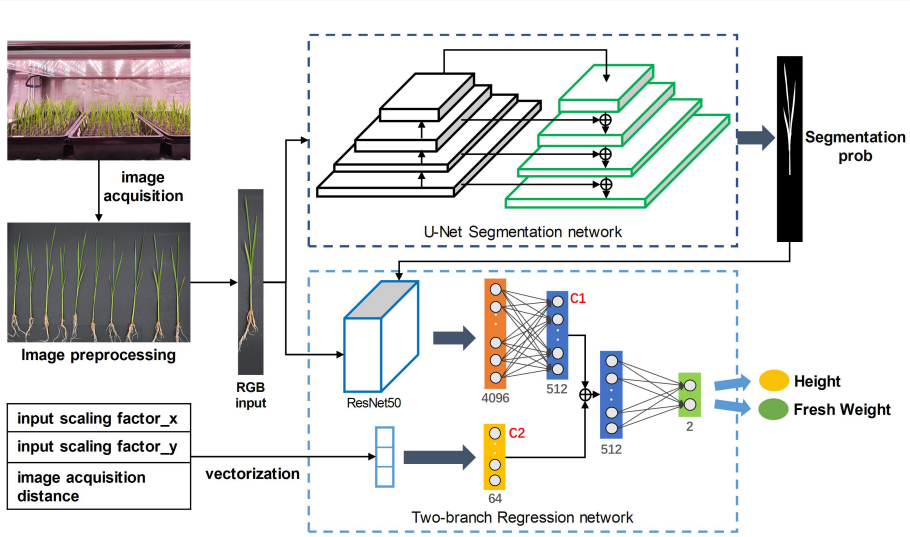


FIGURE 1
Overall structure of the hybrid CNN framework.

criteria for segmentation. Mean absolute error (MAE), normalized root mean square error (NRMSE) and coefficient of determination (R^2) were calculated to evaluate the estimation performance. These metrics are defined as follows:

$$\mathrm{OA} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP},\ \mathrm{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\mathrm{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

$$\mathrm{IoU} = \frac{TP}{FP + TP + FN} \tag{4}$$

where TP is the number of correctly classified seedling pixels, FP is the number of pixels misclassified as seedlings, FN is the number of pixels misclassified as background, TN is the number of correctly classified background pixels.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_i| \tag{5}$$

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2}}{\bar{y}} \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{7}$$

where n is the number of samples, f_i is the *i-th* predicted trait, y_i is the *i-th* ground truth trait, $\bar{y}$is the average of ground truth.

To further evaluate the estimation performance of the proposed model, the classical machine learning classifier RF (Breiman, 2001) and a regression CNN model (RCNN) were adopted to estimate the growth traits of rice seedlings. RF has shown good performance in the estimation of growth traits of crops and fishes (Saberioon and Císař, 2018; Zhang et al., 2020), while RCNN has been reported in estimating the fresh weight of lettuce directly from the input images (Zhang et al., 2020). To build RF classifier, feature extraction was conducted from digital images of rice seedlings. According to the characteristics of seedlings, low-level image features including color features, morphological features and texture features were extracted. Table 1 lists all the features used to build the RF model. Because the RF model itself can evaluate the importance of features, all low-level features were used to fit the RF model in the experiment.

2.5 K-fold cross validation for regression

The regression dataset comprises 504 images, which is a relatively small dataset in the deep learning community. As mentioned in the part of data preprocessing above, 80% of the samples were randomly selected for modeling, and the remaining 20% were used for evaluation. To prevent overfitting, K-fold cross-validation (K=5) was used to build the model on the training set (Stone, 1974). The average of metrics on the test set was taken as the evaluation standard.

3 Results

3.1 Segmentation results of the model

As demonstrated by the accuracy evaluation of the proposed method on the rice seedling segmentation task, the segmentation

TABLE 1 List of image features of rice seedlings.

No.	Feature	Description	Symbols and formulas
1	Average	The average of each color component in five color spaces (RGB, HLS, HSV, CIELab, YCbCr)	Ave
2	Standard deviation	The standard deviation of each color component in 5 color spaces (RGB, HLS, HSV, CIELab, YCbCr)	σ
3	Area	The number of pixels in the seedling area	A
4	Perimeter	The number of pixels of the seedling boundary	P
5	Aspect ratio	Ratio of length to width of minimum rectangle of seedling	$Ar = {}^{L}/_{W}$
6	Ellipse ratio	Ratio of the major axis to the minor axis of the seedling equivalent ellipse	$Er = {}^{a_{ellipse}}/_{b_{ellipse}}$
7	Compactness	Ratio of the diameter of the minimum enclosing circle to the length of the minimum rectangle	$Cpa = {}^{2r}/_{L}$
8	Arc	Ratio of the seedling area to the area of the minimum enclosing circle	$Arc = {}^{A}/_{\pi r^2}$
9	Extent	Ratio of the seedling area to the area of the minimum rectangle	$Ex = {}^{A}/_{L\times W}$
10	Complexity	Ratio of the square of the seeding perimeter to the seedling area	${}^{Cpl=P^2}/_{A}$
11	Contrast	Contrast of the gray level co-occurrence matrix in seedling region	
12	Correlation	Correlation of the gray level co-occurrence matrix in seedling region	
13	Energy	Energy of the gray level co-occurrence matrix in seedling region	
14	Homogeneity	Homogeneity of the gray level co-occurrence matrix in seedling region	

submodel achieved an OA of 0.997, an F1 accuracy of 0.956 and the IoU accuracy of 0.916 (Table 2). Visual interpretation on the test set indicated that the proposed method can distinguish seedling pixels from background pixels well (Figure 2). In conclusion, the model can recognize complete rice seedlings with high accuracy.

3.2 Estimation results of the model

The results of growth-related traits estimation based on the proposed method are shown in Figures 3 and 4, where the figures depict the prediction results in the five-fold cross-validation. Table 3 showed the performance of the proposed model on the test sets for estimating growth-related traits of rice seedlings. The results indicate a strong correlation between the actual measured values of rice seedling growth-related traits and the CNN-based model estimates. In terms of height traits, the regression submodel had a good estimation performance with an average R^2 of 0.980 and an average NRMSE of 2.64%. The results of seedling shoot biomass estimation were slightly worse, with an average R^2 of 0.717 and an average NRMSE of 17.23%.

3.3 Comparison of the results with the conventional estimation methods

The random forest (RF) model was constructed based on the features selected above (Table 1), and the number of trees in the RF model was set as 1000 by grid search. The estimation results of RF classifier were shown in Table 4, Figures 5 and 6. The average R^2 of RF for height estimation results was 0.819, with an average NRMSE of 7.93%. Similar to the results of the CNN model, the estimation performance of biomass traits was lower than that of height traits, with an average R^2 of 0.634 and an average NRMSE of 19.41%.

The estimation results of the RCNN model were shown in Figures 7 and 8. For the two traits of SH and SFW (Table 5), the RCNN model had average R^2 values of 0.688 and 0.492, respectively, and the average NRMSE values were 10.81%, and 23.27%, respectively. The results showed that this regression CNN model struggled in estimating seedling growth-related traits.

4 Discussion

4.1 Comparative analysis of different models

A hybrid segmentation and regression network was built in this study. Compared with existing methods that combine threshold segmentation with deep regression networks, this work adopts a two-stage model to solve segmentation and estimation tasks in an end-to-end manner. In the segmentation stage, a deep segmentation network was used instead of the classic image segmentation algorithms (such as threshold segmentation) mainly for the consideration of computational efficiency. Since the seedling images used in this research were acquired by different sensors, using a data-driven deep segmentation network can automatically learn effective features from images, reducing the difficulty of handcrafted feature design and expert knowledge. In the regression stage, image segmentation prediction is taken as explicit input to help the regression network focus on the seedling pixel region rather than the redundant background pixels. This makes the whole model easier to be optimized on a small dataset. Furthermore, the proposed method is flexible for varying image acquisition distance since the acquisition distance and image scaling factor are taken as inputs to the regression network branches when constructing the model. Meanwhile, the model has good scalability and it can be adapted from the dual task of image segmentation and regression to only perform the segmentation or image-based growth character estimation.

In comparison with the classical machine learning methods, the estimation accuracy of the proposed CNN model for growth-related traits was higher than that of the RF estimator, as shown in Tables 3 and 4. The estimation accuracy of the former was 0.083 higher in R^2 for biomass traits and 0.161 higher in R^2 for height traits, and the estimation results of the CNN model also had lower NRMSE. The results demonstrate the advantages of the CNN model in automatically learning complex features from image data. However, the results of RCNN were somewhat counterintuitive. As can be seen from the tables above, the performance of RCNN for seedling traits estimation was lower than that of the proposed model and RF model. This finding indicates the features extracted by the RCNN are not robust enough to estimate rice seedling traits, which

TABLE 2 Confusion matrix of segmentation test set.

Prediction		
Ground truth	Rice	Background
Rice	3,428,953	165,599
Background	146,830	98,088,319
OA	0.997	
F1	0.956	
IoU	0.916	

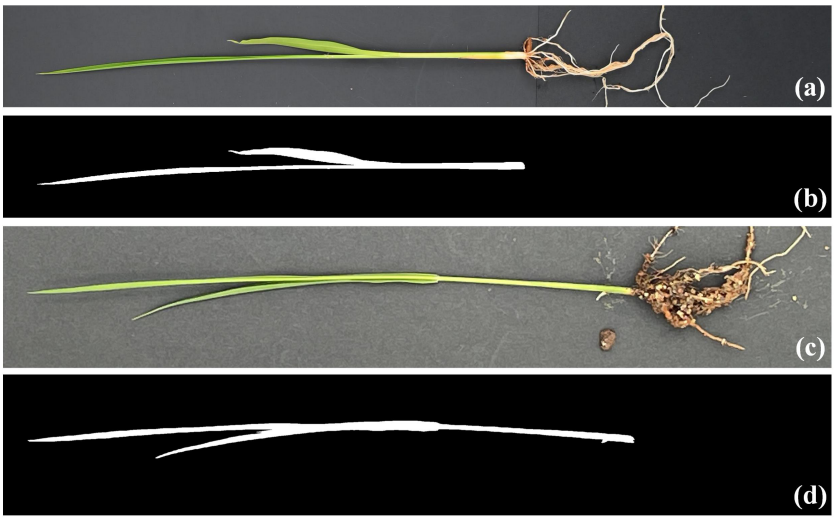


FIGURE 2
Example of results on the segmentation test set **(A, C)** shows the original image taken with a digital camera and a smartphone, respectively, and **(B, D)** shows the corresponding segmentation results.

may be due to the capacity limitation of the five-layer network model. Another possible reason is that the dataset in this paper contains digital images collected by two sensors at different distances, which increases the difficulty of directly estimating growth traits from images.

According to the experimental results, including the proposed method and the other two comparison methods, the performance of seedling height estimation is better than that of biomass estimation. This may be because the information extracted from the plane digital images of a single perspective of rice seedlings can better reflect the characteristics of height traits. However, seedling images from a single perspective still had partial occlusion or hiding, which made the features learned by the model not comprehensive enough for estimating shoot biomass, affecting the estimation accuracy. Therefore, the estimation of height trait by the proposed model is better than that of biomass trait. In our opinion, more training samples may help to improve the performance of biomass estimation.

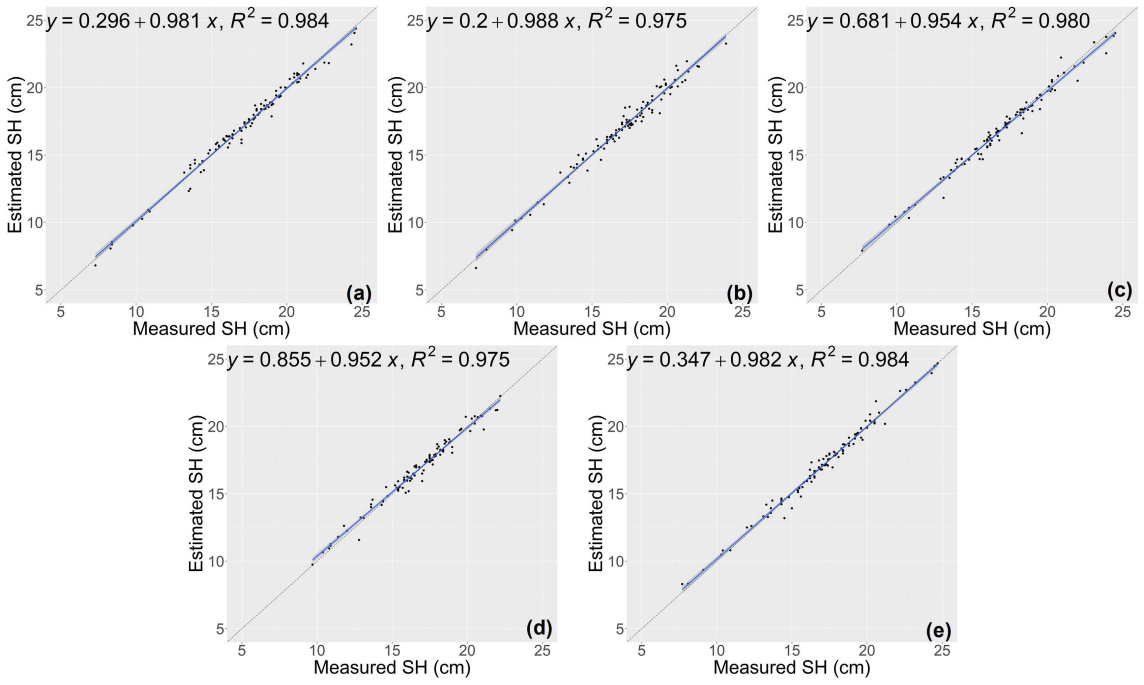


FIGURE 3
Estimation results of shoot height (SH) based on the proposed model **(A–E)** shows the results in the five-fold cross-validation.

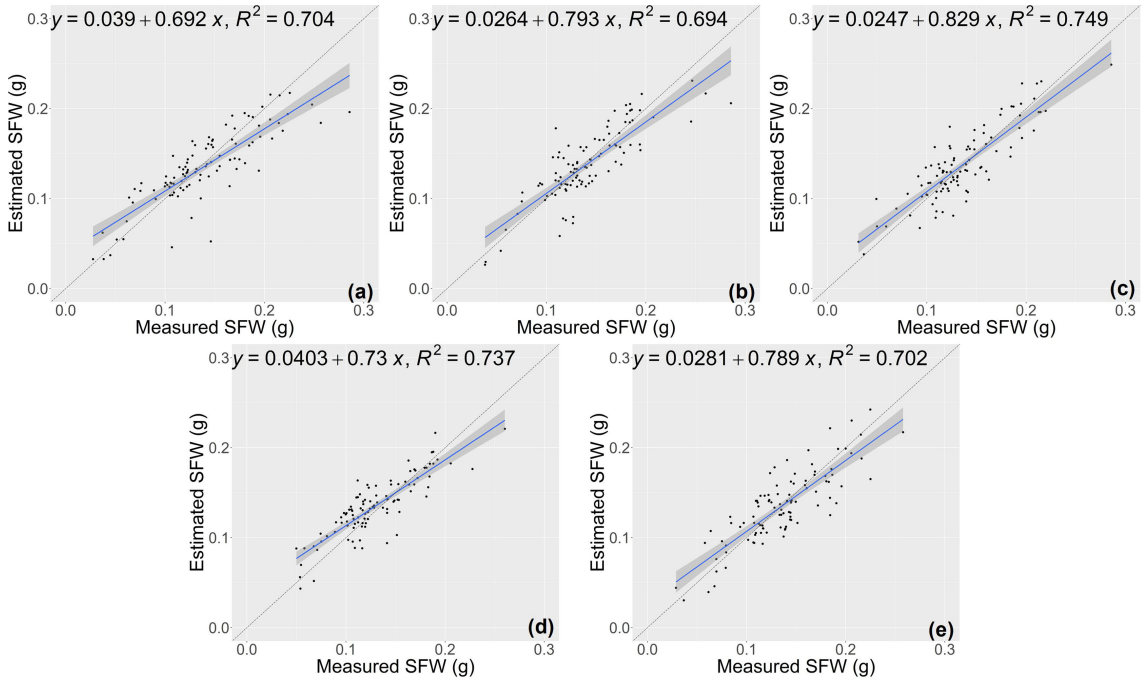


FIGURE 4
Estimation results of shoot fresh weight (SFW) based on the proposed model **(A–E)** shows the results in the five-fold cross-validation.

Nowadays, many reports have introduced image-based approaches to extract phenotypic traits from crop images, with results listed in Table 6. Different from previous works, this study focuses on the growth of rice seedlings in a controlled environment. Specifically, this study explores how CNN-based deep learning techniques can better cope with RGB images acquired under varying acquisition conditions (sensor and acquisition distance). Therefore, acquisition condition-related geometric vectors are considered as branch inputs of the regression network. In addition, unlike some methods that segment images and then extract geometric features, this study integrates image segmentation and image regression into a unified end-to-end framework, prompting the network to automatically learn the implicit representations in segmented images. The experimental results show that the presented approach in this paper is competent for the estimation of growth traits of rice seedlings.

4.2 Parameter analysis of FC layers

In this study, the image acquisition distance and scale factor were taken as the input geometric vector of the branch of the regression CNN, so that the regression network could adapt to the digital images acquired at different shooting distances. In order to explore the effects of the number of neurons in the fully connected layer of regression network, we conducted further comparative experiments. Concretely, the number of neurons in the fully connected layer after the fusion of two paths (denoted as C1) was

TABLE 3 Regression error statistics of the proposed method for growth-related traits.

Number of experiments	SH			SFW		
	MAE(g)↓	NRMSE(%)↓	R^2 ↑	MAE(g)↓	NRMSE(%)↓	R^2 ↑
1	0.338	2.52	0.984	0.020	19.51	0.704
2	0.394	2.92	0.975	0.018	17.05	0.694
3	0.346	2.83	0.980	0.018	16.06	0.749
4	0.323	2.55	0.975	0.017	16.08	0.737
5	0.290	2.35	0.984	0.019	17.44	0.702
Average	0.338	2.64	0.980	0.018	17.23	0.717

TABLE 4 Regression error statistics of the RF method for growth-related traits (n=1000).

Number of experiments	SH			SFW		
	MAE(g)↓	NRMSE(%)↓	R^2 ↑	MAE(g)↓	NRMSE(%)↓	R^2 ↑
1	1.289	9.48	0.774	0.023	22.42	0.613
2	0.913	6.81	0.876	0.020	19.15	0.600
3	1.116	8.09	0.820	0.017	16.43	0.736
4	0.963	7.45	0.797	0.018	18.26	0.657
5	1.063	7.83	0.828	0.022	20.79	0.564
Average	1.069	7.93	0.819	0.020	19.41	0.634

tested, as well as the number of neurons in the fully connected layer adjacent to the branch input (denoted as C2).

Experimental results are used to analyze the value of parameters. From Table 7, it can be seen that the estimation accuracy reaches an optimal level with increasing number of neurons but fluctuates. However, when the number of neurons in the fully connected layer increases to 1024, the estimation accuracy drops due to the excessive number of parameters. Higher values for the number of neurons were not tested due to computing resources limitations. Table 8 shows the experimental results for various settings of C2. Although different values of C2 have different effects on the R^2 of height and fresh weight, the R^2 of both traits outperforms other results by setting C2 = 64. In conclusion, the experimental results prove that 512 and 64 are better choices for parameters C1 and C2, respectively.

4.3 Limitations and future work

Although the results on the test set have proved that the proposed method is accurate and efficient in the seedlings image segmentation and growth traits estimation, the developed framework could still be improved potentially. First, the accuracy of plant height regression is higher than that of fresh weight regression because the images are captured from a single-side view of the rice seedling sample, which requires further mining the information provided by the lateral view to estimate the latter. Moreover, the input of the proposed framework is digital images of monocot plant seedlings, and its applicability to dicotyledonous plant seedlings needs to be further verified.

Future research will continue to collect more seedling images to expand the dataset, and not be limited to a single rice variety. To

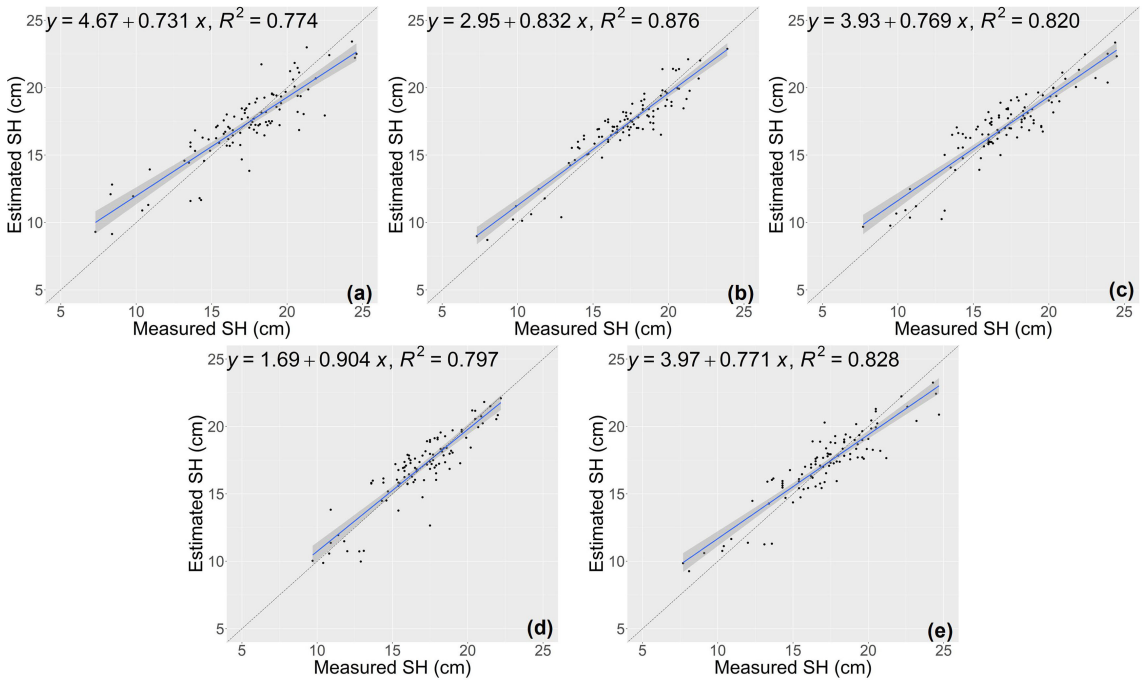


FIGURE 5
Estimation results of shoot height (SH) based on RF model **(A–E)** shows the results in the five-fold cross-validation.

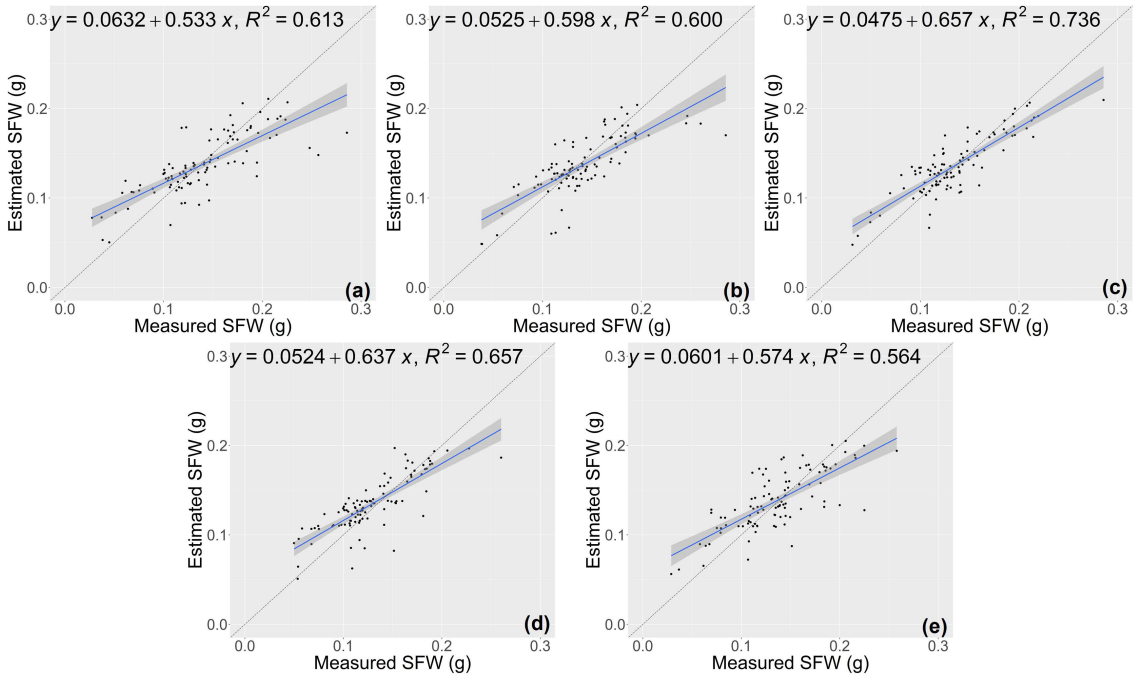


FIGURE 6
Estimation results of shoot fresh weight (SFW) based on RF model **(A–E)** shows the results in the five-fold cross-validation.

further improve the regression prediction accuracy of shoot fresh weight, images from multiple perspectives will be explored as input for the regression model. In addition, it is necessary to develop an automatic acquisition process of multi-view digital images of seedlings in the rice seedling factory, and the method in this study will be improved to adapt to the detection and non-destructive growth monitoring of single rice seedlings in complex backgrounds. Last but not least, the hybrid network is mainly

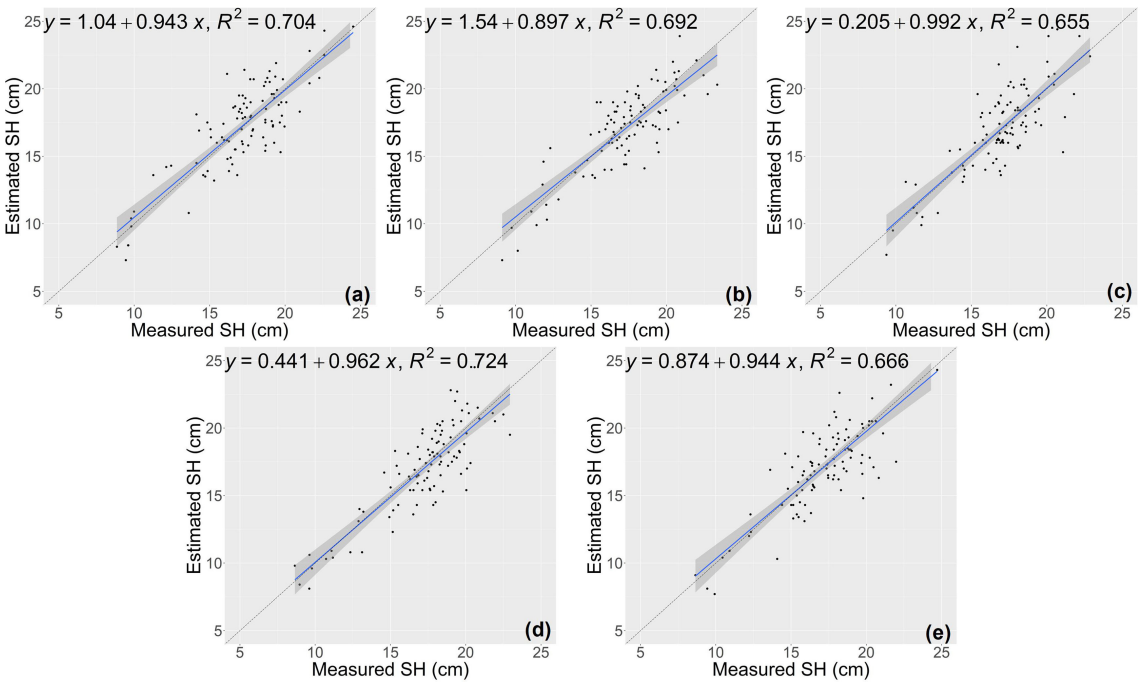


FIGURE 7
Estimation results of shoot height (SH) based on the RCNN model **(A–E)** shows the results in the five-fold cross-validation.

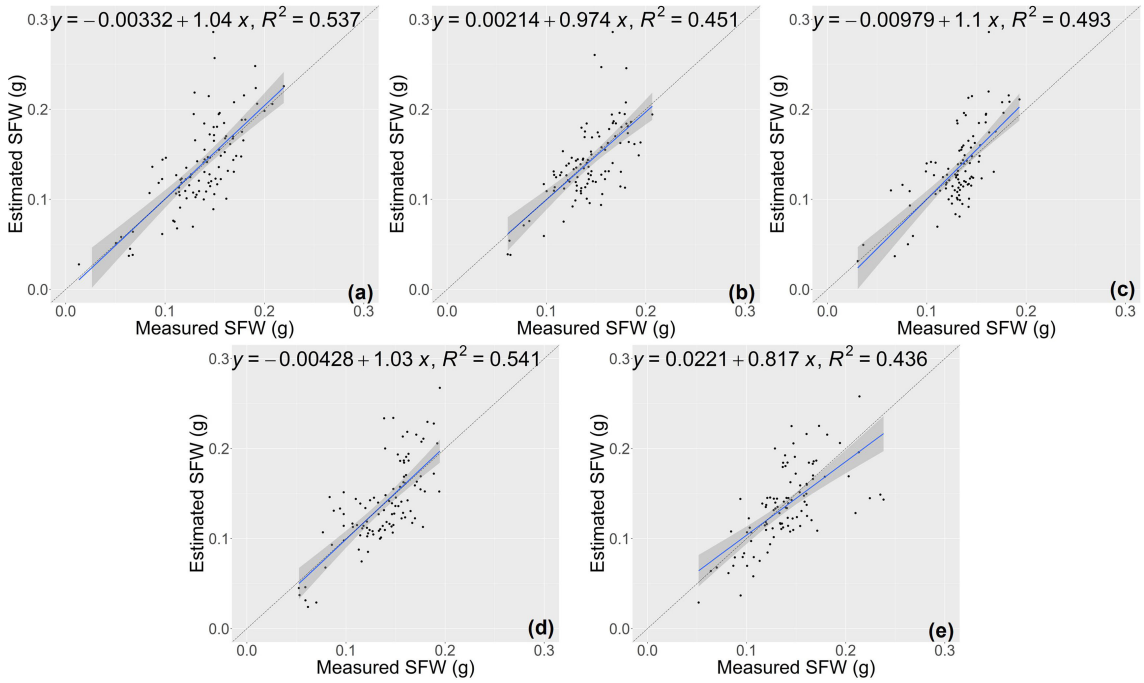


FIGURE 8
Estimation results of shoot fresh weight (SFW) based on the RCNN model **(A–E)** shows the results in the five-fold cross-validation.

composed of UNet and Resnet50, which can be easily deployed on edge computing devices or mobile phones. This means that the presented method is expected to be used for stationary automated phenotyping equipment in plant factories, as well as handheld mobile phenotyping equipment. When combined with mobile devices or integrated with an edge computing platform into vertical seedling factory facilities, this method has great application potential.

5 Conclusion

Rapid acquisition of morphological traits of rice seedlings can help to understand the growth status of rice seedlings, which is the key basis for intelligently controlling the environment of industrial seedlings and making lighting strategies. In this study, a semantic segmentation and growth-related traits estimation method for rice seedlings in a plant factory based on CNN and digital images was proposed, which could facilitate rice seedling growth monitoring. This method supports multiple intelligent terminals such as digital cameras or mobile phones as image acquisition devices and does not need to fix the image acquisition distance, so it is feasible in practical applications. The method was experimentally verified on the rice seedling dataset. The segmentation accuracy of rice seedling achieved an OA of 0.997, an F1 accuracy of 0.95 and an IoU accuracy of 0.91, and the estimated growth-related traits, such as plant height and shoot fresh weight, were also in good agreement with the measured values, with R^2 values of plant height reaching 0.980 and NRMSE reaching 2.64%, R^2 values of shoot fresh weight reaching 0.717 and NRMSE reaching 17.23%. The experimental

TABLE 5 Regression error statistics of the RCNN method for growth-related traits.

Number of experiments	SH			SFW		
	MAE(g)↓	NRMSE(%)↓	R^2 ↑	MAE(g)↓	NRMSE(%)↓	R^2 ↑
1	1.559	10.83	0.704	0.025	24.22	0.537
2	1.442	10.93	0.692	0.023	22.22	0.451
3	1.508	11.01	0.655	0.024	22.60	0.493
4	1.455	10.45	0.724	0.026	23.17	0.541
5	1.419	10.84	0.666	0.025	24.16	0.436
Average	1.477	10.81	0.688	0.025	23.27	0.492

TABLE 6 An overview of existing image-based methods for plant growth traits estimation.

Ref.	Plants	Input data types	Methods	Descriptions
(Chen et al., 2016)	Lettuce	RGB	Image segmentation + Linear regression	Fresh weight regression by geometric features extracted from a stereo-vision system.
(Zhang et al., 2020)	Lettuce	RGB	CNN regression	Multi-objective regression (Fresh weight, dry weight and area) by a five-layer CNN model.
(Lin et al., 2022)	Lettuce	RGB-D	CNN segmentation + CNN regression	Fresh weight regression by a ResNet34 model, fusing RGB-D features and geometric features extracted from segmented images.
(Buxbaum et al., 2022)	Lettuce	RGB-D	CNN regression	Fresh weight regression by a ResNet50 model, fusing RGB-D image features
(Ubbens and Stavness, 2017)	Arabidopsis thaliana	RGB	CNN regression	Age regression by a five-layer CNN model
(Quan et al., 2021)	Weed	RGB-D	CNN regression	Fresh weight regression by a DenseNet201 model, fusing RGB-D image features
This work	Rice seedling	RGB from different sensors	CNN segmentation + CNN regression	Multi-objective regression (Fresh weight and height) by a hybrid CNN framework (UNet + ResNet50), fusing RGB features, segmented images and geometric elements.

TABLE 7 Regression error statistics of the proposed method for different C1 settings.

Parameter setting	SH			SFW		
C1	MAE(g)↓	NRMSE(%)↓	R^2 ↑	MAE(g)↓	NRMSE(%)↓	R^2 ↑
128	0.644	5.36	0.928	0.024	21.04	0.621
256	0.733	5.49	0.925	0.024	21.21	0.615
512	0.338	2.64	0.980	0.018	17.23	0.717
1024	0.788	5.88	0.914	0.024	22.68	0.560

TABLE 8 Regression error statistics of the proposed method for different C2 settings.

Parameter setting	SH			SFW		
C2	MAE(g)↓	NRMSE(%)↓	R^2 ↑	MAE(g)↓	NRMSE(%)↓	R^2 ↑
32	0.525	4.00	0.953	0.020	18.75	0.650
64	0.338	2.64	0.980	0.018	17.23	0.717
128	0.431	3.15	0.975	0.024	21.21	0.614
256	0.346	2.76	0.981	0.022	21.11	0.618

results showed that the proposed method can accurately estimate the growth-related traits of rice seedlings using low-cost and easily accessible RGB digital images.

It can be concluded that the proposed method is a reliable estimation tool for growth-related traits at seedling stage of rice, which has good application potential in seedling growth monitoring. The accurate regression of growth-related traits can further provide support for scientific planting management and selection of varieties at seedling stage. In addition, because the proposed method is based on the common morphological characteristics of crops at seedling stage, it is promising to be used to realize the estimation of growth-related traits of other crops at seedling stage.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ZY and DK conceived and designed the experiments. ZY and XT conducted the experiments, collected the data and wrote the manuscript. MD and XC collected the data and performed the statistical tests. PN provided the technical support. YL, YR, and DK

read and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Key R&D Program of Zhejiang Province (Grant No. 2022C02026 and 2022C02003) and Zhejiang Provincial Basic Public Welfare Research Project of China under Grant No. LGN22C130016.

Acknowledgments

We would like to thank Xueli Wu for the help in preparing plant materials.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ares, G., Ha, B., and Jaeger, S. R. (2021). Consumer attitudes to vertical farming (indoor plant factory with artificial lighting) in China, Singapore, UK, and USA: a multi-method study. *Food Res. Int.* 150, 110811. doi: 10.1016/j.foodres.2021.110811

Borianne, P., Subsol, G., Fallavier, F., Dardou, A., and Audebert, A. (2018). GT-RootS: an integrated software for automated root system measurement from high-throughput phenotyping platform images. *Comput. Electron. Agric.* 150, 328–342. doi: 10.1016/j.compag.2018.05.003

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Buxbaum, N., Lieth, J. H., and Earles, M. (2022). Non-destructive plant biomass monitoring with high spatio-temporal resolution *via* proximal RGB-d imagery and end-to-End deep learning. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.758818

Chen, W.-T., Yeh, Y.-H. F., Liu, T.-Y., and Lin, T.-T. (2016). An automated and continuous plant weight measurement system for plant factory. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00392

Furbank, R. T., and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Gong, L., Du, X., Zhu, K., Lin, C., Lin, K., Wang, T., et al. (2021). Pixel level segmentation of early-stage in-bag rice root for its architecture analysis. *Comput. Electron. Agric.* 186, 106197. doi: 10.1016/j.compag.2021.106197

He, L., Wu, F., Du, X., and Zhang, G. (2022). Cascade-SORT: a robust fruit counting approach using multiple features cascade matching. *Comput. Electron. Agric.* 200, 107223. doi: 10.1016/j.compag.2022.107223

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition". In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. doi: 10.1109/CVPR.2016.90

Hüther, P., Schandry, N., Jandrasits, K., Bezrukov, I., and Becker, C. (2020). ARADEEPOPSIS, an automated workflow for top-view plant phenomics using semantic segmentation of leaf states. *Plant Cell* 32, 3674–3688. doi: 10.1105/tpc.20.00318

Lin, Z., Fu, R., Ren, G., Zhong, R., Ying, Y., and Lin, T. (2022). Automatic monitoring of lettuce fresh weight by multi-modal fusion based deep learning. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.980581

Lu, H., Liu, J., Kerr, P. G., Shao, H., and Wu, Y. (2017). The effect of periphyton on seed germination and seedling growth of rice (Oryza sativa) in paddy area. *Sci. Total Environ.* 578, 74–80. doi: 10.1016/j.scitotenv.2016.07.191

Ma, L. X., Ge, Y. Y., Sun, R. X., Guo, S. L., and Guo, Y. B. (2014). Design and implementation of intelligent control system in rice seedlings factory. *Adv. Mat. Res.* 952, 271–274. doi: 10.4028/www.scientific.net/AMR.952.271

McClements, D. J., Barrangou, R., Hill, C., Kokini, J. L., Lila, M. A., Meyer, A. S., et al. (2021). Building a resilient, sustainable, and healthier food supply through innovation and technology. *Annu. Rev. Food Sci. Technol.* 12, 1–28. doi: 10.1146/annurev-food-092220-030824

Mekawy, A. M. M., Abdelaziz, M. N., and Ueda, A. (2018). Apigenin pretreatment enhances growth and salinity tolerance of rice seedlings. *Plant Physiol. Biochem.* 130, 94–104. doi: 10.1016/j.plaphy.2018.06.036

Mortensen, A. K., Bender, A., Whelan, B., Barbour, M. M., Sukkarieh, S., Karstoft, H., et al. (2018). Segmentation of lettuce in coloured 3D point clouds for fresh weight estimation. *Comput. Electron. Agric.* 154, 373–381. doi: 10.1016/j.compag.2018.09.010

Osco, L. P., de Arruda, M.d. S., Marcato Junior, J., da Silva, N. B., Ramos, A. P. M., Moryia, É.A.S., et al. (2020). A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS J. Photogrammetry Remote Sens.* 160, 97–106. doi: 10.1016/j.isprsjprs.2019.12.010

Perugachi-Diaz, Y., Tomczak, J. M., and Bhulai, S. (2021). Deep learning for white cabbage seedling prediction. *Comput. Electron. Agric.* 184, 106059. doi: 10.1016/j.compag.2021.106059

Quan, L., Feng, H., Lv, Y., Wang, Q., Zhang, C., Liu, J., et al. (2019). Maize seedling detection under different growth stages and complex field environments based on an improved faster r–CNN. *Biosyst. Eng.* 184, 1–23. doi: 10.1016/j.biosystemseng.2019.05.002

Quan, L., Jiang, W., Lou, Z., and Chen, L. (2021). Two-stream dense feature fusion network based on RGB-d data for the real-time prediction of weed aboveground fresh weight in a field environment. *Remote Sens.* 13, 2288. doi: 10.3390/rs13122288

Reyes-Yanes, A., Martinez, P., and Ahmad, R. (2020). Real-time growth rate and fresh weight estimation for little gem romaine lettuce in aquaponic grow beds. *Comput. Electron. Agric.* 179, 105827. doi: 10.1016/j.compag.2020.105827

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: convolutional networks for biomedical image segmentation". In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science*. Cham: Springer International Publishing. pp. 234–241. doi: 10.1007/978-3-319-24574-4_28

Saberioon, M., and Císař, P. (2018). Automated within tank fish mass estimation using infrared reflection system. *Comput. Electron. Agric.* 150, 484–492. doi: 10.1016/j.compag.2018.05.025

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B. Methodol.* 36, 111–133. doi: 10.1111/j.2517-6161.1974.tb00994.x

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01190

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O., and Schurr, U. (2020). Phenotyping: new windows into the plant for breeders. *Annu. Rev. Plant Biol.* 71, 689–712. doi: 10.1146/annurev-arplant-042916-041124

Yu, Z., Cao, Z., Wu, X., Bai, X., Qin, Y., Zhuo, W., et al. (2013). Automatic image-based detection technology for two critical growth stages of maize: emergence and three-leaf stage. *Agric. For. Meteorology* 174–175, 65–84. doi: 10.1016/j.agrformet.2013.02.011

Zhang, J., Bai, Z., Huang, J., Hussain, S., Zhao, F., Zhu, C., et al. (2019). Biochar alleviated the salt stress of induced saline paddy soil and improved the biochemical characteristics of rice seedlings differing in salt tolerance. *Soil Tillage Res.* 195, 104372. doi: 10.1016/j.still.2019.104372

Zhang, Lu, Wang, J., and Duan, Q. (2020). Estimation for fish mass using image analysis and neural network. *Comput. Electron. Agric.* 173, 105439. doi: 10.1016/j.compag.2020.105439

Zhang, L., Xu, Z., Xu, D., Ma, J., Chen, Y., and Fu, Z. (2020). Growth monitoring of greenhouse lettuce based on a convolutional neural network. *Hortic. Res.* 7, 124. doi: 10.1038/s41438-020-00345-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 June 2023
DOI 10.3389/fpls.2023.1169709

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
United States Department of Agriculture,
United States

REVIEWED BY
Myongkyoon Yang,
Jeonbuk National University,
Republic of Korea
Chan-Woo Jeon,
AGMO Inc., Republic of Korea

*CORRESPONDENCE
Dae-Hyun Jung
jeoguss@gmail.com

RECEIVED 20 February 2023
ACCEPTED 28 April 2023
PUBLISHED 02 June 2023

CITATION
Jung D-H, Kim H-Y, Won JH and Park SH (2023) Development of a classification model for *Cynanchum wilfordii* and *Cynanchum auriculatum* using convolutional neural network and local interpretable model-agnostic explanation technology.
Front. Plant Sci. 14:1169709.
doi: 10.3389/fpls.2023.1169709

COPYRIGHT
© 2023 Jung, Kim, Won and Park. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of a classification model for *Cynanchum wilfordii* and *Cynanchum auriculatum* using convolutional neural network and local interpretable model-agnostic explanation technology

Dae-Hyun Jung[1]*, Ho-Youn Kim[2,3], Jae Hee Won[4] and Soo Hyun Park[2]

[1]Department of Smart Farm Science, Kyung Hee University, Yongin-si, Republic of Korea, [2]Smart Farm Research Center, Korea Institute of Science and Technology (KIST), Gangneung-si, Republic of Korea, [3]Division of Bio-Medical Science and Technology, Korea Institute of Science and Technology (KIST) School, University of Science and Technology (UST), Daejeon, Republic of Korea, [4]Quality Certification Center, National Development Institute of Korean Medicine (NIKOM), Daegu, Republic of Korea

Cynanchum wilfordii is a perennial tuberous root in the Asclepiadaceae family that has long been used medicinally. Although *C. wilfordii* is distinct in origin and content from *Cynancum auriculatum*, a genus of the same species, it is difficult for the public to recognize because the ripe fruit and root are remarkably similar. In this study, images were collected to categorize *C. wilfordii* and *C. auriculatum*, which were then processed and input into a deep-learning classification model to corroborate the results. By obtaining 200 photographs of each of the two cross sections of each medicinal material, approximately 800 images were employed, and approximately 3200 images were used to construct a deep-learning classification model *via* image augmentation. For the classification, the structures of Inception-ResNet and VGGnet-19 among convolutional neural network (CNN) models were used, with Inception-ResNet outperforming VGGnet-19 in terms of performance and learning speed. The validation set confirmed a strong classification performance of approximately 0.862. Furthermore, explanatory properties were added to the deep-learning model using local interpretable model-agnostic explanation (LIME), and the suitability of the LIME domain was assessed using cross-validation in both situations. Thus, artificial intelligence may be used as an auxiliary metric in the sensory evaluation of medicinal materials in future, owing to its explanatory ability.

KEYWORDS

explainable artificial intelligence, image classification, Incpetion-ResNet, sensory evaluation analysis, image classification algorithm

1 Introduction

Northeast Asian countries have long employed medical plants, such as herbal medicines, in the private sector to treat chronic ailments, and the industry is constantly developing. Functional compounds extracted from medicinal plants, in particular, are being employed as components of health functional meals, and efficacy studies as pharmaceutical substances are being actively done. Thus, the value of their use is quite high (Jiang et al., 2011). As the need for medical plants grows, so do related industrial achievements. However, the requirement for periodic standardization and stability difficulties before production, distribution, and sales management is growing (Han et al., 2016). This is a fundamental issue to address in maintaining the equivalence of efficacy and quality control of pharmaceuticals employing accurate medicinal ingredients, and ensuring such quality is a critical aspect in standardizing and modernizing oriental medicine products for entry into the global market.

Cynanchum wilfordii is a perennial tuberous plant of the *Asclepiadaceae* family that grows in Korea, China, and Japan. It has traditionally been used for nutrition, tonic, blood, and renal health. Among the components, gagaminine has been shown to inhibit hepatic aldehyde oxidase, and cynandione A has been shown to inhibit nerve cell damage (Ryuk et al., 2014). Comparing the chemical composition of the two herbs, Caudatin-2,6-dideoxy-3-O-methy- β-D-cymaropyranoside and caudatin isolated from C. auriculatum Royle ex Wight showed high antitumor activity against SMMC-7721 cells and inhibited H22 tumor growth *in vivo* (Peng et al., 2008). Meanwhile, C. wilfordii roots yielded nine new C21 steroidal glycosides, cynawilfosides A–I, and 12 known compounds(Li et al., 2016). C. *wilfordii* is not related to *C. auriculatum* in origin or content. However, recognizing it is challenging for the public owing to the similarity between the ripe fruit and seed. *C. auriculatum*, in particular, was designated as a toxic plant. However, two separate raw materials are blended and placed into the market before ensuring raw material stability, which might lead to consumer distrust of the mixed product (Ryu et al., 2018). It is particularly challenging to detect whether or not it is blended since it is processed in the form of powder or extract.

The traditional method of identifying medicines through morphological characteristics such as appearance, color, taste, and smell is widely used in the field because it has the advantage of being simple and direct confirmation. However, its fundamental flaw is dependent on the discriminator's expertise and subjectivity (Jiang et al., 2011; Ryu et al., 2018). Various methods have been developed to compensate for these disadvantages. Internal morphology identification using microscopy and anatomical techniques, physicochemical identification that analyzes and compares components, and gene identification through unique genetic information search through genome analysis (Kim et al., 2013). A method using DNA markers, which are not affected by external factors, has recently been actively studied for major crops to identify plant species accurately and quickly. Unlike morphological characteristics such as the shape and size of plants, the method using DNA markers can distinguish plant species without being affected by the external environment. Furthermore, no limit exists on the number of markers that can be used. Therefore, an accurate discrimination is possible (Sato-Masumoto et al., 2017). However, it is necessary to secure the DNA nucleotide sequence of all medicinal materials, and there is a limit to immediately distinguishing raw materials in the field.

The spread of smartphones and the development of deep-learning-based image recognition technology have laid the foundation for easily recognizing various objects and acquiring information through cameras (Sun and Qian, 2016; Afonso et al., 2020). The field of plant recognition is a good use case that can apply the advancement of these technologies, and various technological studies are being conducted. Furthermore, services based on this have become an environment that is easily accessible through mobile devices (Grinblat et al., 2016; Saleem et al., 2019; Xi and Panoutsos, 2020; Jung et al., 2022). The massive amount of ImageNet data has had a profound impact on the development of machine learning technology in the field of image recognition vision. In particular, deep-learning technology, which has developed existing neural network technology, has shown rapid performance improvement. It has been applied to almost all fields that require artificial intelligence (AI), such as vision and natural language processing, and research is still being actively conducted. Plant recognition technology mainly utilizes convolutional neural networks (CNN) among deep-learning technologies. Recognizing the type of plant captured by the camera is one of the primary purposes of plant recognition. Because of this, plants become the main subject, and usability is sufficiently increased by classification alone. Dyrmann et al. (2016) developed a CNN model with 86.2% classification accuracy for 22 weed and crop species by learning 10,413 images, including 22 weed and crop species. Lin et al. (2019) applied a CNN-based deep-learning model to classify Powdery Mildew in cucumber and reported a classification accuracy of about 96.08%.

Deep-learning technology is widely applied in the plant recognition field with the development of deep-learning technology. However, there are not many cases of collecting and analyzing large amounts of images of medicinal plants. In particular, the most important point in applying image-based deep-learning classification technology is to enhance understanding of which patterns and parts of the sample were identified to predict the results for plants such as Cynanchum wilfordii, which are difficult to distinguish due to the large number of similar varieties with the naked eye.

Because these medicinal samples are easily jumbled during the processing and distribution stages, they always necessitate the services of a sensory evaluation expert (Chong and Cros, 2004). Numerous images are required in this sensory evaluation area to translate morphological features into data, and it is critical to locate the feature locations. However, due to the black-box nature of neural network-based modeling, it is nearly hard to interpret, making model prediction challenging to understand. Sun and Qian (2016) used CNN technology for herbal medicine image recognition to classify 95 categories of medicinal herbs from 5523 images. As a result, it reported 71% recognition accuracy and 53% retrieval accuracy on average when learning. This demonstrates that there is still a tremendous possibility for improvement if we contribute to learning by spotting similarities among therapeutic ingredients.

Consequently, Explainable Artificial Intelligence (XAI) technology, which can provide AI interpretation, is gaining popularity. By building more explainable models while maintaining high levels of performance, XAI enables analysts to understand and trust AI (Apostolopoulos et al., 2022; Başaran, 2022). The goal of XAI was defined in three ways: The first is to reduce the model's complexity, the second is to improve the predictability of model predictions, and the third is to employ AI models for decision-making. One of the existing detection models, a rule-based model, was proposed in early XAI as a model that conducts detection based on rules that analysts can understand. Zhen et al. suggested a CNN learning structure with an interpretability-oriented layer in the form of fuzzy logic-based rules. The LRP (Layer-wise Relevance BackPropagation) model was proposed as a method of backtracking the NN (Neural Network) model's outcomes and calculating the contribution of input data to individual features (Bach et al., 2015). Local Interpretable Model-Agnostic Explanations (LIME) is an explanation technique for local models that focuses on training local surrogate models to explain prediction outcomes (Ribeiro et al., 2016; Tulio Ribeiro et al., 2016). LIME is frequently used for image-based active site prediction. It looks for 'superpixels' with the highest expressiveness for binary vectors or class outputs that signal the presence or absence of continuous pathways. Several research cases have been published, particularly those that learn characteristics and enhance confidence in CNN algorithms.

CNN models capable of distinguishing *C. wilfordii* and *C. auriculatum* are used in this study to verify the viability of image-based categorization among herbal medications. Furthermore, this study aims to contribute to the advancement of deep-learning recognition models by employing LIME, which can express morphological features using explainable AI technology.

2 Materials and methods

2.1 Acquisition of image data and preprocessing method

C. wilfordii and *C. auriculatum* belong to the same Asclepiadaceae family and are physically quite similar because they have hypertrophied roots. Tuberous roots, in particular, are generally seen in horizontally or vertically cut forms. The entire length is approximately 7 to 12 cm, and the diameter is approximately 0.5 to 1.5 cm. Figure 1A depicts the overall shape of *C. wilfordii* roots. It is commonly supplied on the market in the shape of B and C in Figure 1, and samples cut in this manner are difficult to match with *C. auriculatum*. Figure 1B exhibits the X-Y axis with the long side of the root as the reference axis, while Figure 1C illustrates the Y-Z axis. The truncated B and C forms of *C. wilfordii* and *C. auriculatum* are to be recognized individually in this study. As a result, two cut sections were photographed, yielding 800 images, 200 of each of the four classes.

The amount of training data is significant for constructing a model in AI learning. The data augmentation method is primarily used when obtaining a large number of image data is difficult. Data augmentation was done on the captured images in this study using the generally used image augmentation approach depicted in Figure 2 below. In Figure 2, A is the original image, and B-G are images augmented using a commonly used image processing technique.

In this study, 100 images were applied to each augmentation method by selecting original images using a random function to avoid overfitting through a specific technique in the six preprocessing techniques. In the following, we investigate how the learning and verification performance is affected by whether or not image augmentation is conducted for the dataset before applying it to the AI model.

2.2 Development of CNN classification model

2.2.1 Applied CNN model

In RGB image object recognition and classification, the 2D CNN model outperforms existing image processing approaches (Rawat and Wang, 2017). Two models with good performance and applications in various industries were chosen from among the most frequently reported model structures in 2D image categorization and utilized as comparison groups in this study. Moreover, the same inputs and outputs were employed as in the previous 2D CNN proposal. VGGnet-19 is the first 2D-CNN algorithm identified for comparison. VGGNet is a model proposed by Simonyan and Zisserman (2014) in the 2014 ILSVRC, which came in second place after GoogLeNet. However, it is used more often since it is structurally simpler than GoogLeNet, thus, easier to understand and test. Figure 3 illustrates the structure

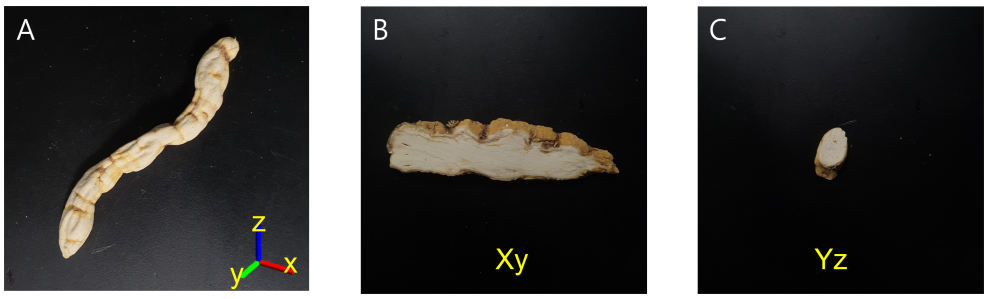


FIGURE 1
A sample of the root zone of *C. wilfordii* **(A)**, horizontal section (Xy) **(B)**, vertical section (Yz) **(C)**.

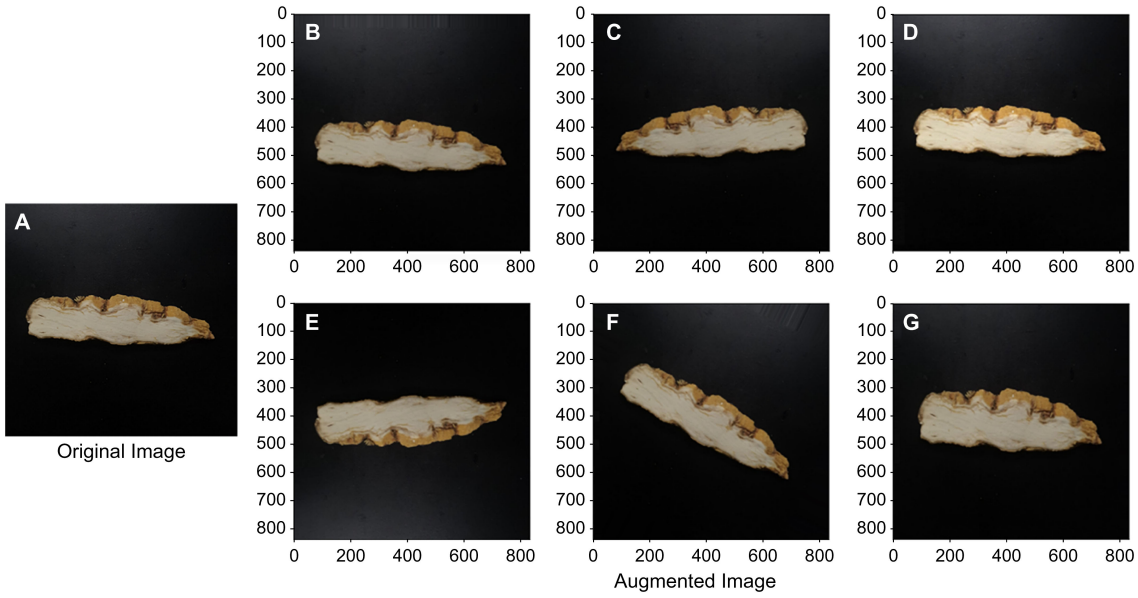


FIGURE 2
Original image of the horizontal section sample of *C. wilfordii* **(A)** and augmented images **(B)** Vertical shift; **(C)** Horizontal Flip; **(D)** Brightness; **(E)** Vertical Flip; **(F)** Rotation; **(G)** Zooming.

of VGGnet-19. Layers are learned from the input on the left to the softmax on the right of the diagram. In general, it consists of a conv layer, which serves as the core of the CNN structure, and a pooling layer, which reduces data space. Before categorization, the process's ultimate outcome is flattened one-dimensionally through a fully linked layer. The softmax function is then used for categorization. In the case of VGGnet, the filter of the conv layer was set to 3 × 3 with stride = 1, and the ReLU function was utilized as the activation function.

Inception-ResNet, proposed by Szegedy et al. (2017), is the second CNN classification model used. Inception-ResNet was designed to successfully broaden and deepen the Inception neural network. Furthermore, better learning speed was reported by adopting a simpler and more uniform structure than Inception-v4 and more Inception modules. Figure 4 depicts the entire structure of Inception-ResNet.

The structure of the stem used has a conv-net structure commonly seen in general CNN structures, as shown in Figure 5A below. This is because the inception effect is not significant at the beginning of the layer. The biggest feature of Inception CNN is that the matrix operation itself is processed densely while connecting the Conv layer sparsely. As shown in Figure 5B, a residual connection structure is configured to implement a function that calculates by combining the result of the previous layer with the result of the current layer. This has significantly affected the improving learning speed. The Resnet-A layer has a 35 x 35 grid, and the Resnet-B layer has a 17 x 17 grid. The modules to reduce this are consecutively composed of Reduction-A and B.

2.3 Local interpretable model-agnostic explanation for XAI

LIME is a method that specifically implements the local surrogate model. Surrogate models learn to "approximate" the



FIGURE 3
The structure of VGGnet-19 used to develop the image classification model.

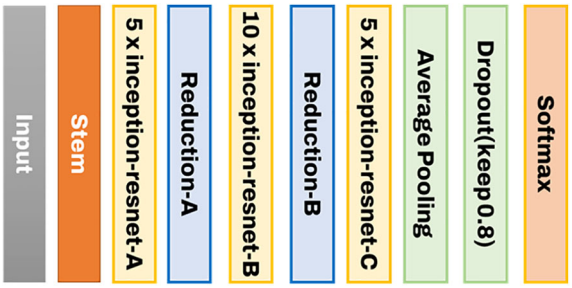


FIGURE 4
The structure of Inception-ResNet used to develop the image classification model.

predictions of the underlying black box model. However, we focus on learning a local surrogate model to explain individual predictions rather than learning a global surrogate model. A good approach to using LIME is to forget about the original dataset and see why the model returned a particular result. Therefore, LIME tests what changes occur to model predictions when the input data is modified and creates a new dataset consisting of perturbed samples and corresponding predicted values of the black box model. Based on this, an interpretable model weighted by the proximity of the sampled observation to the observation (instance) of interest is learned. A local surrogate model has interpretability constraints can be expressed as follows (Ribeiro et al., 2016):

$$\textbf{explanation}(x) = \textbf{argmin}_{g \in G}\, L(f, g, \pi_x) + \Omega(g) \quad \ldots[1]$$

Here g denotes a model that can explain the observed value x and minimizes the loss function L. While keeping the complexity of the model Ω(g) lower, the loss measures how close the predictions of the explanatory model are to those of the original model f. The loss function L here is expressed as follows (Ribeiro et al., 2016):

$$L(f, g, \pi_x) = \sum_{z, z' \in Z} \pi_x(z)(f(z) - g(z'))^2 \quad \ldots[2]$$

Here G is a set of possible explanations, including a general linear regression model, and π_x is the measure of proximity indicating how close to the observed value (x) that we consider for the explanation.

Images' data format is a cluster of pixels, and it is difficult to determine how significantly a single pixel contributes to class determination. Individual pixel changes have little effect on the model's predictions, and image transformation is accomplished through the superpixel segmentation and masking procedure. Superpixel is a method for image classification that collects and groups perceptually important pixels (Wang et al., 2017). Adjacent pixels with comparable qualities (e.g., color, brightness) are clustered together to form a big pixel or superpixel. It signifies that the image will be processed in units of superpixels rather than units of pixels. When we look at an image, we look at related parts together rather than each pixel separately. As a result, it can be demonstrated that processing images in superpixel units is a more natural (human-like) way (Achanta et al., 2012). Thus, superpixel and XAI technology were coupled to provide an auxiliary metric for sensory evaluation. The image was segmented using Simple linear iterative clustering (SLIC), one of the superpixel segmentation algorithms. SLIC is an algorithm applying k-means for superpixel generation and has two key differences. 1) Reduce the number of distance computations in the optimization by limiting the search space to an area equivalent to the superpixel size. This becomes linear in the number of pixels N and can lower the complexity irrespective of the number of superpixels. 2) A weighted distance measurement maintains consistency while using color and geographical proximity. In this image classification and XAI application, around 250 pieces were processed using the SLIC algorithm on 512 × 512 video images (Figure 6).

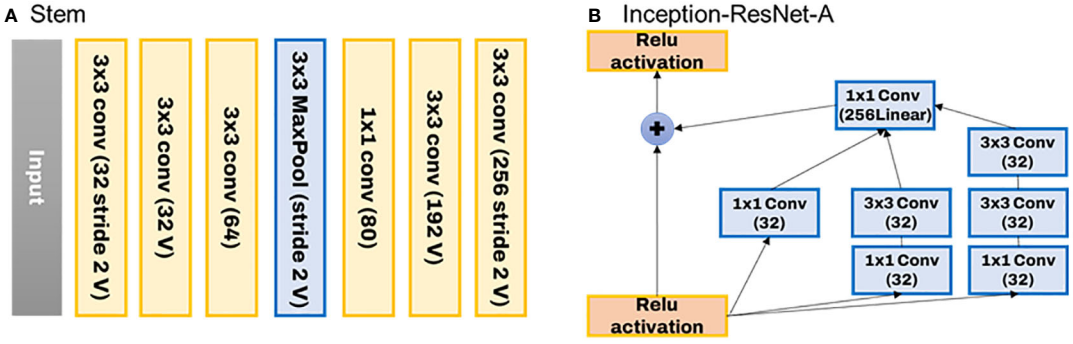


FIGURE 5
Stem structure used in Inception-ResNet **(A)** and ResNet-A layer of the residual connection structure **(B)**.

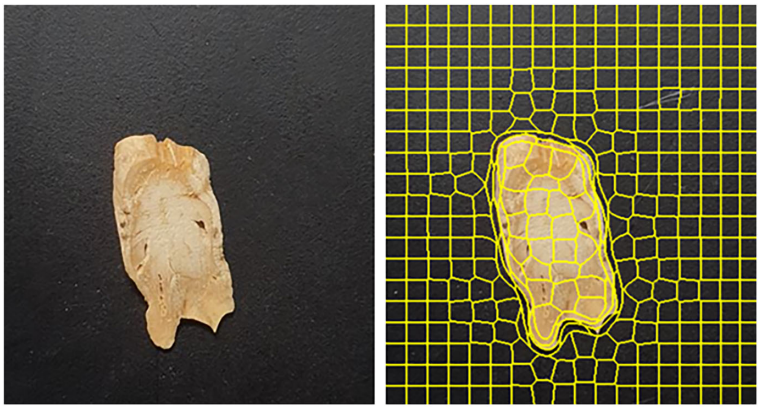


FIGURE 6
Top view of the cut *C. wilfordii* (right) and image segmented by the SLIC algorithm for superpixels (left).

2.4 Model performance evaluation method

The developed model is a classification model for differentiating *C. wilfordii* and *C. auriculatum*, and its performance was evaluated based on accuracy, F1 score, and the receiver operating characteristic (ROC) curve. Accuracy is a metric that takes into account the situation in which the model infers two classification labels and predicts true as true and false as false, and is expressed as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \quad \ldots[3]$$

True Positive (TP): Predict the answer that is actually true as true (correct answer)

False Positive (FP): Predict the answer that is actually false as true (wrong answer)

False Negative (FN): Predict the answer that is actually true as false (wrong answer)

True Negative (TN): Predict the answer that is actually false as false (correct answer)

The F1 score is one of the statistics that define the classification accuracy and recall rate, which are combined into a single statistic. Here, the harmonic average was determined, not the standard average. Its purpose is to ensure that the F1 score has a similar low value as precision and recall, which are close to 0. The equation for the F1 score is as follows:

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precisioin} \cdot \text{recall}}{\text{precision}+\text{recall}} \quad \ldots[4]$$

Figure 6 depicts the classification of the training data set used to compare the performance of the CNN model used to create the *C. wilfordii* and *C. auriculatum* classification models. Models 1-1 and 1-2 apply unprocessed video images without a separate light reduction procedure. Models 2-1 and 2-2, which will be contrasted with this, eliminate background light reflection. Except for the sample area, the background was transformed to picture RGB (0,0,0) or black. The classification goals of Model 1-1 and Model 1-2 were assigned distinct labels to classify the two shortened forms of *C. wilfordii* and *C. auriculatum*. Models 1-2 and 2-2 defined both portions as a single medicinal product label and then classified each variant (Figure 7). Moreover, whether a region that could serve as a sensory metric could be detected was determined by applying LIME to each model.

The CNN Inception-ResNet and VGGnet models were constructed using the Tensorflow and Keras libraries in Python 3.8. The GPU used to calculate the model was NVIDIA RTX 3090.

3 Results

3.1 Comparison of performance between CNN Inception-ResNet and VGGnet

The collected images of *C. wilfordii* and *C. auriculatum* were divided into four classes based on the cut section, and learning and verification were performed. Furthermore, the learning and verification performances were compared using two CNN models. Through the augmentation of the original 800 images, around 3200 images were utilized. The learning-to-verification ratio was 8:2, and the data was split. Figure 8 compares the outcomes of data models 1-1 and 2-1. In the performance of the two CNN models, the CNN Inception-ResNet structure clearly exhibited faster learning speed and higher classification accuracy than VGGnet-19. The learning accuracy of Model 1-1's Inception-ResNet was 0.835, while its verification accuracy was 0.812. The learning accuracy of VGGnet was 0.776, while the validation accuracy was 0.710. The classification accuracy of model 2-1 with light reflection removal was 0.821 at 0.861 verification accuracy in Inception-ResNet training and 0.701 at 0.712 verification accuracy in VGGnet training. Figure 9 is a confusion matrix displaying the classification accuracy of Inception-four ResNet's classes. The average separation accuracy of the xy-cut sections in collected samples of *C. wilfordii* and *C. auriculatum* was 0.86, but the yz-cut sections had a lower accuracy of 0.76.

With the implementation of the data augmentation method, the accuracy in the test set showed a noteworthy improvement of 0.1 to

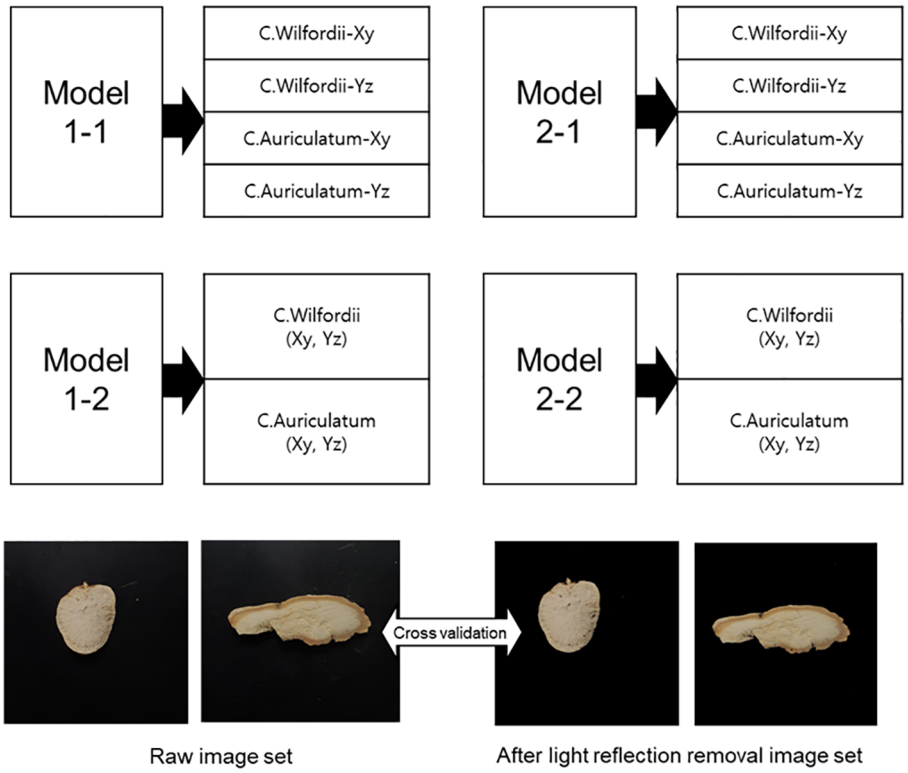


FIGURE 7
Training data groups to compare the performance of CNN models applied to develop the *C. wilfordii* and *C. auriculatum* classification models.

0.15. This indicates that there was a significant boost in learning performance, even with a limited number of samples, while also mitigating the potential risk of overfitting.

3.2 Result of removing light reflection of an image through the application of LIME

Models 2-1 and 2-2 corroborated the model classification outcome by specifying two cut sections as one class. Table 1 summarizes the results. Models 1-1 and 1-2, which split and classified cut sections into classes, showed higher classification performance in general. In particular, the accuracy of VGGnet-19 increased considerably from 0.71 to 0.80. The performance of Inception-ResNet improved slightly as well, and the validation set classification accuracy in Model 2-2 was around 0.862.

Table 2 summarizes the performance cross-validation results of Model 1-2, a classification model developed using raw images, and Model 2-2, which was developed using images with light removal processing to extract sample regions. The performance of the model trained with raw images was confirmed to be much lower when categorizing images with light reflection removed. The accuracies of 0.641 and 0.671 were validated as the performance of the two CNN models, which failed to show significant performance in classifying the two classes.

These results validated the local active area identified by the LIME algorithm as the cause. Figures 10, 11 depict the trained model as a raw image and the image with the background's reflected light removed, respectively. They display the outcome of discovering an explanation component through LIME analysis of the trained model. Figure 10 depicts not only the cross-section of the sample but also the descriptive local area inside the region of reflected light in the backdrop. Conversely, Figure 11 confirmed the outcome of locating the explanatory component in the sample area.

4 Discussion

Medicinal plants have similar species that can be readily jumbled in distribution, and distinguishing them requires a very high level of sensory evaluation experience. The medicinal plant species *C. wilfordii* and *C. auriculatum* exhibit similar morphologies, and they are commonly intermingled in the distribution process. These kinds are exceedingly tough to find professionals that can recognize them in the field. As a result, disruptive molecular analysis, such as physicochemical component analysis or RNA analysis, can provide confidence in their categorization accuracy (Han et al., 2016). However, this study demonstrated that it is possible to differentiate between *C. wilfordii* and *C. auriculatum* efficiently using AI image classification technology, which is rapidly evolving.

In particular, the architectures of CNN models Inception-ResNet and VGGnet-19 were used for classification. Inception-ResNet demonstrated improved results in terms of performance

10.3389/fpls.2023.1169709

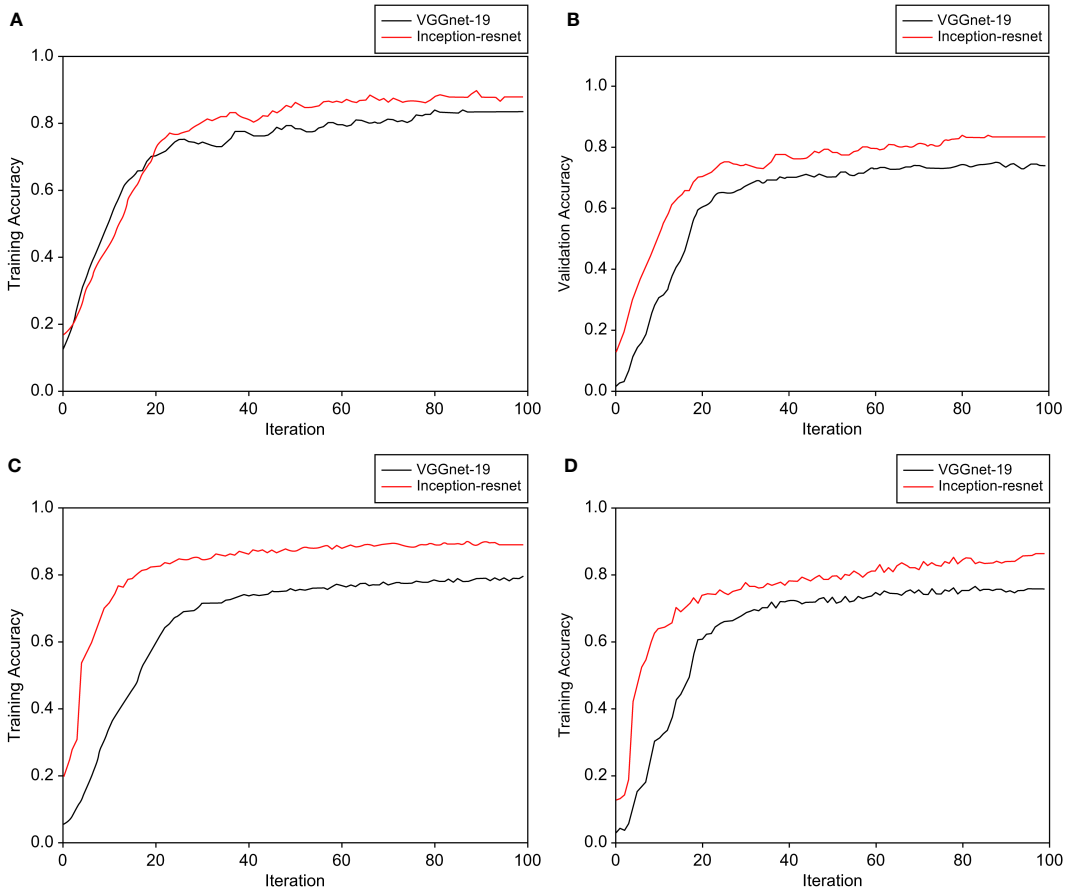


FIGURE 8
Comparison of the performance of the CNN model applied to develop a classification model for *C. wilfordii* and *C. auriculatum*: **(A)** Classification result of raw image learning model, **(B)** Verification result of the developed raw image-learning model, **(C)** Classification result of the learning model after light removal treatment, and **(D)** Verification result of the developed light-cancellation image set model.

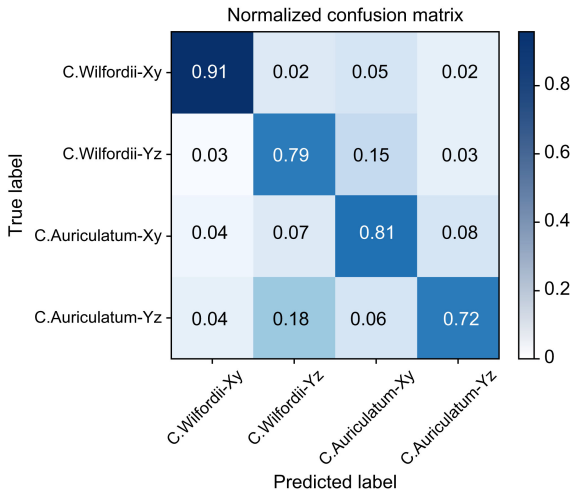


FIGURE 9
Four-label classification confusion matrix of the CNN Inception-ResNet.

and learning rate. Regarding two algorithms, the average processing speed for the same training data (approximately 2400 images) in a single iteration was 12.45 ± 1.25s and 26.22 ± 2.76s, respectively, based on the GPU specifications used. A strong classification performance of around 0.862 was confirmed in the validation set. The two medications are mostly available in sliced form. When 4 classes were identified through 2 representative cross-sectional examples, the highest verification result was roughly 0.835. The xy-axis truncated shape demonstrated higher classification accuracy. In fact, the model appears to have limitations if the model is inferred or given in a crushed or another form while the sample is not cut. In this study, images were captured at a consistent height and position of the sample in the top view, where the cross section is most visible. Thus, the classification performance loss is to be expected when testing the classification performance of images from multiple camera angles. This appears to be solvable by acquiring more diversified and numerous images and applying them for training.

LIME, an AI explainability technique, was employed in this study to investigate whether it may be used as an auxiliary metric for sensory evaluation by marking the explanatory component among cross-sections of categorized medicinal materials. Figure 11 confirmed that the technique was explanatory in the sample's local area and suitably adaptable. Light reflection on the background is

10.3389/fpls.2023.1169709

TABLE 1 Validation set accuracy and F1 scores of classification cases 2-1 and 2-2.

	Inception-ResNet		VGGnet-19	
	Accuracy	F1 score	Accuracy	F1 score
Model 1-2	0.855	0.834	0.811	0.806
Model 2-2	0.862	0.851	0.805	0.801

TABLE 2 Cross-validation accuracy and F1 score results of classification cases 2-1 and 2-2 based on the developed CNN model.

	Inception-ResNet		VGGnet-19	
	Accuracy	F1 score	Accuracy	F1 score
Model 1-2 to Model 2-2	0.641	0.560	0.671	0.622
Model 2-2 to Model 1-2	0.821	0.798	0.742	0.712

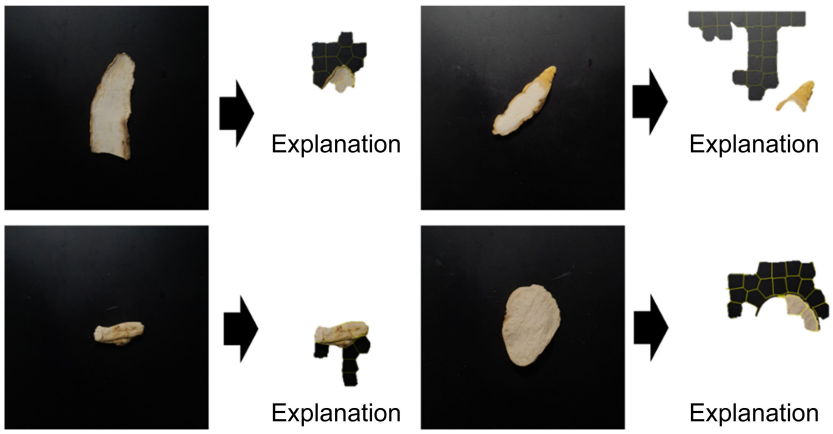


FIGURE 10
Result of finding explanation parts through LIME analysis for the Inception-ResNet model trained by raw images.

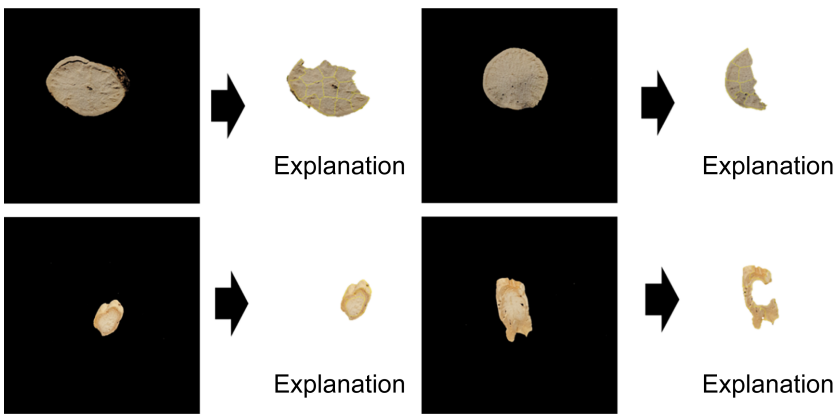


FIGURE 11
Result of finding the explanation part of the Inception-ResNet model through LIME analysis for the Inception-ResNet model trained by images from which the reflected light on the background has been removed.

evident to the naked eye in the raw image data of the collected image. LIME yielded sample results with an explanatory ability to model training in this domain. This appears to be due to the model detecting the impact of the surrounding environment during image collecting; this is expected to be validated later through LIME analysis of several images taken in different situations. Only the sample had model explainability in the area where the background light was deleted. It was also established that the surface of the sample, rather than the center, is the area primarily stimulated between the *C. wilfordii* and *C. auriculatum* samples; this can also be perceived as the overall shape of the cross-sectional contour.

Explainable AI refers to the technology that provides interpretable forms of the prediction results generated by machine learning models. This technology helps to understand and analyze the prediction results of the model, making the model's prediction results more reliable. The image recognition and classification technology in video and images has already been extensively researched, but it is difficult to judge whether the learning intentions of the model, which is developed as a black box, match. By utilizing explainable AI technology, it becomes possible to understand how a machine learning model makes predictions, enabling the judgment of whether the model's learning intentions are consistent. Furthermore, the classification technology of C. wilfordii and C. auriculatum could be used as an auxiliary means for determining herbal medicine in the distribution field if it is connected to a web platform, and it is highly likely that it can be used as a means of providing accurate information to customers by using devices that can acquire images, such as smartphones. This technology can increase the reliability of herbal medicine sales and help customers use herbal medicine correctly.

5 Conclusions

Images were collected in this work to categorize *C. wilfordii* and *C. auriculatum*, which were then processed and put into a deep-learning classification model to corroborate the results. For image classification, the architectures of Inception-ResNet and VGGnet-19 among CNN models were employed for classification. Inception-ResNet demonstrated improved results in terms of performance and learning rate. The validation set confirmed a strong classification performance of around 0.862. LIME was also used to add explanatory characteristics to the deep-learning model. In both situations, the appropriateness of the LIME area was determined using cross-validation. As a result, the raw image was confirmed to be activated in the light reflection area in the surrounding background. When eliminated, the created model's accuracy declined dramatically from 0.855 to 0.641. The second case model designed to choose the sample region, on the other hand, maintained an accuracy of 0.8 or higher even after cross-validation. This Explainable AI has the potential to be employed as an auxiliary metric in the sensory evaluation of therapeutic compounds in future.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, D-HJ and SP. Methodology, D-HJ. Validation, H-YK and JW. Formal analysis, JW. Data curation, D-HJ. Writing—original draft preparation, D-HJ. Writing—review and editing H-YK and JW, Visualization, SP. Supervision, D-HJ. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by a grant from Kyung Hee University in 2022 (KHU- 20222211).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 2274–2282. doi: 10.1109/TPAMI.2012.120

Afonso, M., Fonteijn, H., Fiorentin, F. S., Lensink, D., Mooij, M., Faber, N., et al. (2020). Tomato fruit detection and counting in greenhouses using deep learning. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.571299

Apostolopoulos, I. D., Athanasoula, I., Tzani, M., and Groumpos, P. P. (2022). An explainable deep learning framework for detecting and localising smoke and fire incidents: evaluation of grad-CAM++ and LIME. *Mach. Learn. Knowl. Extr.* 4, 1124–1135. doi: 10.3390/make4040057

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., and Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS One* 10, e0130140. doi: 10.1371/journal.pone.0130140

Başaran, E. (2022). Classification of white blood cells with SVM by selecting SqueezeNet and LIME properties by mRMR method. *Signal. Image. Video. Process.* 16, 1821–1829. doi: 10.1007/s11760-022-02141-2

Chong, P. S. T., and Cros, D. P. (2004). Technology literature review: quantitative sensory testing. *Muscle Nerve Off. J. Am. Assoc. Electrodiagn. Med.* 29, 734–747. doi: 10.1002/mus.20053

Dyrmann, M., Karstoft, H., and Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. doi: 10.1016/j.biosystemseng.2016.08.024

Grinblat, G. L., Uzal, L. C., Larese, M. G., and Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 127, 418–424. doi: 10.1016/j.compag.2016.07.003

Han, E.-H., Cho, K., Goo, Y., Kim, M., Shin, Y.-W., Kim, Y.-H., et al. (2016). Development of molecular markers, based on chloroplast and ribosomal DNA regions, to discriminate three popular medicinal plant species, cynanchum wilfordii, cynanchum auriculatum, and polygonum multiflorum. *Mol. Biol. Rep.* 43, 323–332. doi: 10.1007/s11033-016-3959-1

Jiang, Y., Choi, H. G., Li, Y., Park, Y. M., Lee, J. H., Kim, D. H., et al. (2011). Chemical constituents of cynanchum wilfordii and the chemotaxonomy of two species of the family asclepiadaceae, c. wilfordii and c. auriculatum. *Arch. Pharm. Res.* 34, 2021–2027. doi: 10.1007/s12272-011-1203-z

Jung, D.-H., Kim, C. Y., Lee, T. S., and Park, S. H. (2022). Depth image conversion model based on CycleGAN for growing tomato truss identification. *Plant Methods* 18, 1–15. doi: 10.1186/s13007-022-00911-0

Kim, M., Wang, H., Kim, Y., Sathiyamoorthy, S., Kwon, W., and Yang, D. (2013). Molecular authentication by multiplex-PCR of three similar medicinal plant species: cynanchum wilfordii, cynanchum auriculatum and polygonum multiflorum (Fallopia multiflorum). *J. Med. Plants Res.* 7, 2584–2589. doi: 10.5897/JMPR12.1299

Li, J.-L., Gao, Z.-B., and Zhao, W.-M. (2016). Identification and evaluation of antiepileptic activity of C21 steroidal glycosides from the roots of cynanchum wilfordii. *J. Nat. Prod.* 79, 89–97. doi: 10.1021/acs.jnatprod.5b00766

Lin, K., Gong, L., Huang, Y., Liu, C., and Pan, J. (2019). Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Front. Plant Sci.* 10, 155. doi: 10.3389/fpls.2019.00155

Peng, Y., Li, Y., Liu, X., Zhang, J., and Duan, J. (2008). Antitumor activity of c-21 steroidal glycosides from cynanchum auriculatum royle ex Wight. *Phytomedicine* 15, 1016–1020. doi: 10.1016/j.phymed.2008.02.021

Rawat, W., and Wang, Z. (2017). Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* 29, 2352–2449. doi: 10.1162/neco_a_00990

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv. Prepr. arXiv1606.05386*, 91–95. doi: 10.48550/arXiv.1606.05386

Ryu, H. J., Kim, A. K., Kim, S. D., Jung, S. J., Jang, J. I., Lee, H. J., et al. (2018). Genetic authentication of cynanchi wilfordii radix and cynanchi auriculati radix by using conventional-PCR and real-time PCR. *Korean. J. Pharmacogn.* 49, 55–64. doi: 10.1007/s11760-022-02141-2

Ryuk, J. A., Lee, H. W., Ju, Y. S., and Ko, B. S. (2014). Monitoring and identification of cynanchum wilfordii and cynanchum auriculatum by using molecular markers and real-time polymerase chain reaction. *J. Korean. Soc Appl. Biol. Chem.* 57, 245–251. doi: 10.1007/s13765-013-4248-5

Saleem, M. H., Potgieter, J., and Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants* 8. doi: 10.3390/plants8110468

Sato-Masumoto, N., Uchikura, T., Sugiwaki, H., Yoshimura, M., Masada, S., Atsumi, T., et al. (2017). Survey on the original plant species of crude drugs distributed as cynanchi wilfordii radix and its related crude drugs in the korean and chinese markets. *Biol. Pharm. Bull.* 40, 1693–1699. doi: 10.1248/bpb.b17-00226

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv. Prepr. arXiv1409.1556*, 1–14. doi: 10.48550/arXiv.1409.1556

Sun, X., and Qian, H. (2016). Chinese Herbal medicine image recognition and retrieval by convolutional neural network. *PloS One* 11, e0156327. doi: 10.1371/journal.pone.0156327

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. in. *Proc. AAAI. Conf. Artif. Intelligence*, 4278–4284. doi: 10.1609/aaai.v31i1.11231, 4278-4284

Tulio Ribeiro, M., Singh, S., and Guestrin, C. (2016). " why should I trust you?": explaining the predictions of any classifier. *ArXiv. e-prints.* arXiv-1602. doi: 10.1145/2939672.2939778

Wang, M., Liu, X., Gao, Y., Ma, X., and Soomro, N. Q. (2017). Superpixel segmentation: a benchmark. *Signal Process. Image. Commun.* 56, 28–39. doi: 10.1016/j.image.2017.04.007

Xi, Z., and Panoutsos, G. (2020). "Interpretable convolutional neural networks using a rule-based framework for classification," in *Intelligent systems: theory, research and innovation in applications* (Springer), 1–24.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 July 2023
DOI 10.3389/fpls.2023.1133505

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Rajesh Yakkundimath,
Visvesvaraya Technological University, India
Disha Sushant Wankhede,
Vishwakarma Institute of Technology, India
Satheesh Kumar Perepu,
Ericsson, Sweden

*CORRESPONDENCE
Jianwei Qin
jianwei.qin@usda.gov

RECEIVED 29 December 2022
ACCEPTED 07 June 2023
PUBLISHED 04 July 2023

CITATION
Qin J, Monje O, Nugent MR, Finn JR, O'Rourke AE, Wilson KD, Fritsche RF, Baek I, Chan DE and Kim MS (2023) A hyperspectral plant health monitoring system for space crop production. *Front. Plant Sci.* 14:1133505. doi: 10.3389/fpls.2023.1133505

COPYRIGHT
© 2023 Qin, Monje, Nugent, Finn, O'Rourke, Wilson, Fritsche, Baek, Chan and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A hyperspectral plant health monitoring system for space crop production

Jianwei Qin[1]*, Oscar Monje[2], Matthew R. Nugent[2], Joshua R. Finn[2], Aubrie E. O'Rourke[3], Kristine D. Wilson[3], Ralph F. Fritsche[3], Insuck Baek[1], Diane E. Chan[1] and Moon S. Kim[1]

[1]Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD, United States, [2]Amentum, NASA Kennedy Space Center, Merritt Island, FL, United States, [3]Exploration Research and Technology, NASA Kennedy Space Center, Merritt Island, FL, United States

Compact and automated sensing systems are needed to monitor plant health for NASA's controlled-environment space crop production. A new hyperspectral system was designed for early detection of plant stresses using both reflectance and fluorescence imaging in visible and near-infrared (VNIR) wavelength range (400–1000 nm). The prototype system mainly includes two LED line lights providing VNIR broadband and UV-A (365 nm) light for reflectance and fluorescence measurement, respectively, a line-scan hyperspectral camera, and a linear motorized stage with a travel range of 80 cm. In an overhead sensor-to-sample arrangement, the stage translates the lights and camera over the plants to acquire reflectance and fluorescence images in sequence during one cycle of line-scan imaging. System software was developed using LabVIEW to realize hardware parameterization, data transfer, and automated imaging functions. The imaging unit was installed in a plant growth chamber at NASA Kennedy Space Center for health monitoring studies for pick-and-eat salad crops. A preliminary experiment was conducted to detect plant drought stress for twelve Dragoon lettuce samples, of which half were well-watered and half were under-watered while growing. A machine learning method using an optimized discriminant classifier based on VNIR reflectance spectra generated classification accuracies over 90% for the first four days of the stress treatment, showing great potential for early detection of the drought stress on lettuce leaves before any visible symptoms and size differences were evident. The system is promising to provide useful information for optimization of growth environment and early mitigation of stresses in space crop production.

KEYWORDS

hyperspectral imaging, reflectance, fluorescence, machine learning, plant stress

1 Introduction

In fresh produce production systems to be deployed in future NASA spacecraft, monitoring plant growth and health during crop development from seedling to harvest is needed to ensure the food safety and security of pick-and-eat salad crops consumed by astronauts during cis-lunar, lunar, and Martian missions (Monje et al., 2019). Currently, plant monitoring in growth chambers onboard the International Space Station (e.g., NASA's Veggie and Advanced Plant Habitat) is conducted by estimating plant growth rates based on photographic analysis of daily increments in leaf area. This limited approach cannot detect plant stresses, nutrient deficiencies, and diseases, which usually develop days before any visible leaf changes are observed (Massa et al., 2016; Zeidler et al., 2019). There is a need for novel sensing techniques that can monitor plant health before visible symptoms appear. Compact and automated sensing systems that require minimal crew intervention are preferred for better fit in volume-limited plant growth chambers. Early stress detection ensures the food safety of the crops produced for human consumption.

Nondestructive optical sensing methods and imaging-based technologies have rapidly progressed and been adopted for plant phenotyping (Zhao et al., 2019; Yang et al., 2020). High-throughput phenotyping (HTP) platforms have been developed by public and private sectors for use under both natural field conditions and controlled environments in various applications. The field-based HTP platforms may use aerial-based or ground-based approaches for crop phenotyping. The aerial-based systems typically carry miniaturized and lightweight airborne sensors on unmanned aerial vehicles (UAVs), such as multi-rotor, fixed wing, and flying-wing UAVs as well as helicopters and blimps (Yang et al., 2017). The flight heights and speeds of UAVs can enable crop canopy imaging at field level to cover an entire plot within minutes. The ground-based systems have been built based on manually pushed carts (White & Conley, 2013), self-propelled tractors (Jiang et al., 2018), and overhead systems supported by gantry (Virlet et al., 2017) or cable (Bai et al., 2019). These systems can implement multiple sensing modalities for long-exposure measurements at plant level owing to their high sensor payloads and battery capacities. Crop phenotyping can also be carried out at leaf level using portable sensing devices such as a handheld hyperspectral imager (Wang et al., 2020). The measurements of individual leaves can be combined with geo-location information to create geo-referenced data for mapping plant health in the field. On the other hand, the HTP platforms for controlled-environment agriculture have been developed mainly for use in greenhouses and growth chambers, and can collect high quality phenotypic data from plants with low variability due to relatively uniform environments. Such collection is usually limited and difficult under natural field conditions. Large-scale commercial phenotyping systems designed for the greenhouses usually use plant-to-sensor approaches for automated inspection of plants on conveyor belts viewed from the side or overhead with multiple integrated imaging sensors (Ge et al., 2016). For the growth chambers, the phenotyping systems are generally configured in a sensor-to-plant setup to inspect the plant samples from above (Lien et al., 2019). The sensors can also be installed on robotic arms (e.g., eye-in-hand cameras) to conduct indoor phenotyping for individual plant samples (Bao et al., 2018).

Various imaging modalities have been developed for evaluating plant characteristics, such as growth, stress, and disease. Examples include color imaging for morphology and geometry inspection, near-infrared imaging for leaf water content assessment, fluorescence imaging for chlorophyll content evaluation, hyperspectral imaging for stress monitoring, thermal imaging for leaf temperature measurement, and 3D imaging for shoot and canopy structure profiling (Li et al., 2014). Hyperspectral imaging can simultaneously obtain both spectral and spatial information from a target, which makes it a powerful tool for many food and agricultural applications (Qin et al., 2017). Under controlled indoor environments, close-range hyperspectral reflectance imaging commonly uses halogen lamps for broadband illumination of plants and has been applied to plant phenotyping (Mishra et al., 2017). A variety of hyperspectral image analysis methods have shown promising results for detecting early onset of plant stress and disease (Lowe et al., 2017). Fluorescence imaging is another useful technique for plant monitoring applications, such as evaluation of plant resistance to pathogens (Rousseau et al., 2013) and detection of plant diseases (Pérez-Bueno et al., 2016). Recently, 3D imaging techniques have been combined with close-range hyperspectral imaging to reduce geometry effects of various plant structures on the spectral and spatial data, such as using a line laser scanner based on triangulation principle (Behmann et al., 2016) and a depth sensor based on time-of-flight principle (Huang et al., 2018). Integrating multimodal imaging modalities generally can enhance the sensing capabilities for plant phenotyping (Jiang et al., 2018; Bai et al., 2019; Nguyen et al., 2023). However, few sensing systems have been reported on combining reflectance and fluorescence imaging techniques into one system for *in situ* plant monitoring applications.

Traditional hyperspectral imaging systems based on separated spectrographs and cameras are bulky and heavy, difficult to move frequently or to fit into small spaces. Existing commercial greenhouse phenotyping systems are generally too large as well. Also, the high-power halogen lights commonly used for hyperspectral reflectance imaging generate a lot of heat, which is not ideal for illuminating tender plants at close range. Recently, all-in-one small hyperspectral cameras have been introduced to the market, which made it possible to develop compact imaging systems. Examples include a handheld hyperspectral camera with an integrated scanner for plant disease detection (Behmann et al., 2018), a snapshot hyperspectral camera on a microscope for bacteria identification in dairy products (Unger et al., 2022), and a miniature line-scan hyperspectral camera on a UAV for maize phenotyping (Nguyen et al., 2023). With the advancement of the LED technologies, broadband LED lights can provide an alternative for the traditional halogen lights for the reflectance measurement, which is particularly important when sample heating is a concern, such as in biomedical applications (Stergar et al., 2022). To develop next-generation hyperspectral imaging systems able to autonomously monitor plant health and food safety in future manned space missions, an interagency agreement has been established between USDA Agricultural Research Service (ARS)

and NASA Kennedy Space Center (KSC) to leverage their complementary areas of expertise—ARS sensing technology development and KSC space crop production. As a first step of this collaborative research, this study aimed to develop a compact and automated hyperspectral prototype system for installation in a KSC plant growth chamber to conduct imaging experiments on pick-and-eat salad crops grown under controlled environments. The system was designed to collect both hyperspectral reflectance and fluorescence images within a single imaging cycle using broadband and UV-A LED lights. Specific objectives of this methodology paper were to (1) present the system design, development, software, calibration, operation, and hyperspectral image processing methods for plant samples and (2) demonstrate the system's performance and capability with an example application for detecting drought stress for lettuce.

2 Materials and methods

2.1 Hyperspectral plant health monitoring system

2.1.1 System development

The hyperspectral imaging system developed for automated health inspection of plants grown in a controlled-environment chamber is schematically illustrated in Figure 1. Since the system is designed for the confined space of the growth chamber to perform frequent imaging passes of stationary plants, all major hardware components, including lighting, camera, and translation stage, must be compact and lightweight for an overhead imaging setup. This sensor-to-sample arrangement differs from our previously developed hyperspectral systems that generally image moving samples using stationary lights and cameras (sample-to-sensor) (Kim et al., 2011). The system uses two LED line lights to illuminate the plant samples for hyperspectral image acquisition —one to provide visible and near-infrared (VNIR) light for reflectance imaging and, separately, the other to provide ultraviolet-A (UV-A) excitation light for fluorescence imaging. Specifically, LEDs at five VNIR wavelengths (428, 650, 810, 850, and 915 nm) and one UV-A wavelength (365 nm) are used in the two lights (UX-CLL509-C428C650IR810IR850IR915-24Z-US2 and UX-CLL509-UV365-24Z-US2, Metaphase Technologies, Bristol, PA, USA). Intensities of the LEDs at the six wavelengths can be independently adjusted by two digital dimming controllers (three channels each). Each light is mounted using two pivot joints at an angle of approximately 10° from the vertical position and has a full rod focal lens to create a narrow and structured beam (approximately 54 cm long and 1.5 cm wide on sample surface) with high intensity and concentrated light for a narrow, linear field of view. Each of the two line lights are enclosed in aluminum housings with minimum heat generation during rapid image acquisition.

A compact line-scan hyperspectral camera (microHSI 410, Corning Specialty Materials, Keene, NH, USA) is used to measure light reflectance/fluorescence from the plant samples in the VNIR wavelength range (400–1000 nm). A miniature solid block Offner imaging spectrograph and a CMOS focal plane array detector (12-bit and 1936×1216 pixels) are integrated in a small-form-factor package for the all-in-one hyperspectral camera. A wide-angle low-distortion lens with 5 mm focal length (Edmund Optics, Barrington, NJ, USA) allows imaging coverage of the full width of the growth chamber (53.3 cm) within the height of the chamber. A long-pass

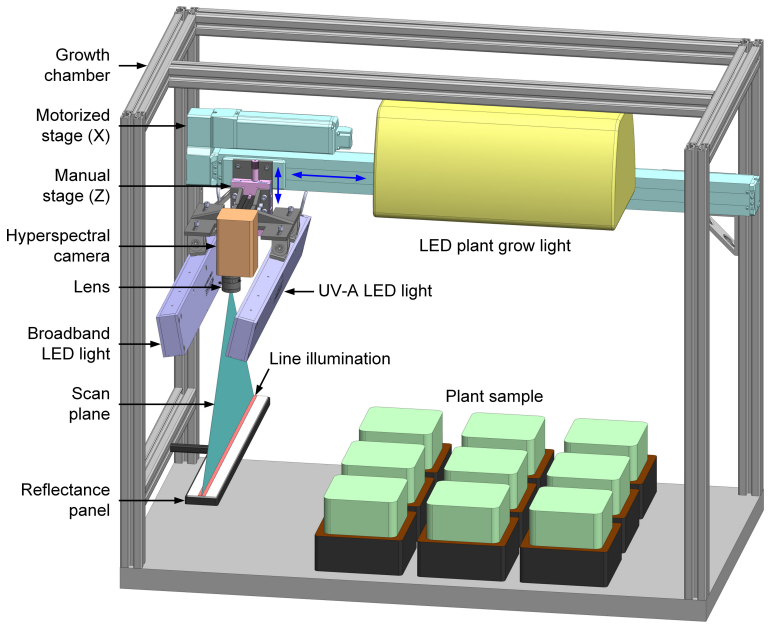


FIGURE 1
A compact and automated hyperspectral reflectance and fluorescence imaging system for plant health monitoring in a controlled-environment growth chamber.

(>400 nm) gelatin filter (Wratten 2A, Kodak, Rochester, NY, USA) in front of the lens removes the UV-A excitation source peak at 365 nm and thus second-order effects around 730 nm. The lights and camera are on a manual translation stage (Thorlabs, Newton, NJ, USA) to enable 5 cm vertical adjustment for imaging plants of different heights. The manual stage is mounted on a linear motorized stage with a stroke of 80 cm (Intelligent Actuator, Los Angeles, CA, USA) that translates the lights and camera for overhead line-scan image acquisition. A reflectance standard panel with a custom size of 50×5 cm^2 (Labsphere, North Sutton, NH, USA) is mounted under the camera's origin position at one end of the chamber for flat-field correction to the reflectance images of the plants. When imaging is not actively taking place, a ceiling-mounted LED plant grow light (RX30, Heliospectra, Gothenburg, Sweden) provides continuous, simulated full-spectrum sunlight for photosynthesis in the plants. The growth chamber equipped with the imaging system is placed in a dark room to avoid the influence of ambient light on both plant growth and imaging.

2.1.2 System software

Imaging system software was developed using LabVIEW (v2017, National Instruments, Austin, TX, USA) in the Microsoft Windows 10 operating system on a laptop computer to provide a user-friendly graphic interface (Figure 2). Software development kits (SDKs) from the hardware manufacturers were used in the LabVIEW programming environment to communicate with major hardware components, including the VNIR and UV-A LED lights, the LED plant grow light, the hyperspectral camera, and the motorized translation stage. Functions from both the SDKs and LabVIEW were used to implement hardware parameterization and data transfer tasks, such as User Datagram Protocol (UDP) for LED light control, Universal Serial Bus (USB) for camera control, LabVIEW Vision Development Module (VDM) for image display and processing, and serial communication for stage movement control. Image acquisition can be started manually at any time or automatically using a timed imaging function (e.g., once a day at 9 AM). During image acquisition, a pair of reflectance and fluorescence images along with an original spectrum and a spatial profile are displayed and updated line by line to show the scan progress in real time. After each measurement, the reflectance and fluorescence images collected from the same scene are saved separately in a standard format of band interleaved by line (BIL) with timestamps appended to the filenames. The saved images can be processed and analyzed offline using in-house programs developed by MATLAB (R2022a, MathWorks, Natick, MA, USA).

2.1.3 System operation

When an imaging cycle is initiated manually or automatically, the LED plant grow light is turned off to eliminate interference to the hyperspectral image acquisition. The VNIR line light is then turned on for 30 s to stabilize the LED output at five wavelengths. Then, the motorized translation stage begins moving towards the far end of the growth chamber. As the stage moves, the hyperspectral camera continuously collects line-scan reflectance signals while passing over the reflectance standard and then the plant samples below. When the sensing unit reaches the far end of the chamber, the reflectance image acquisition is completed and the VNIR light is turned off. Then, the UV-A line light is turned on and the camera begins continuous collection of line-scan fluorescence signals as the stage reverses movement back toward the origin position. When the stage reaches its original starting position, the UV-A light is turned off, completing one full imaging cycle that produces a pair of hyperspectral images, one reflectance and one fluorescence, for the same scene of the plant samples. Finally, the LED plant grow light is turned back on to continue providing simulated sunlight to the plants. In addition to the continuous moving mode, the system can also conduct incremental step-by-step hyperspectral scanning (i.e., stop-and-go mode), which

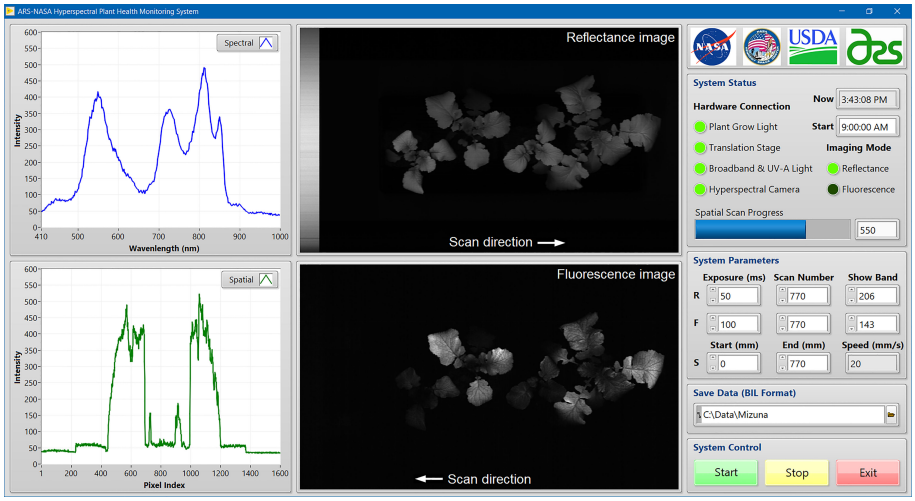


FIGURE 2
In-house developed LabVIEW software for system control and hyperspectral image acquisition from plant samples.

generally does not need image registration to align the reflectance and fluorescence images of the same scene when identical step size is used for both imaging modes.

2.1.4 System calibrations

Spectral calibration was conducted for the hyperspectral camera using five standard pencil calibration lamps, including argon, krypton, neon, xenon, and mercury-neon, to map pixel indices to wavelengths based on a linear regression model (Figure 3). For a total of 1216 pixels along the spectral dimension of the detector, it was found that pixel indices of 401–700 corresponded to the wavelength range of 408–1001 nm with an interval of 1.98 nm. Hence, only 300 pixels in the VNIR region are collected for the spectral acquisition. On the other hand, for spatial calibration of the system using a lens with 5 mm focal length and a working distance of 214 mm, the length of the instantaneous field of view (IFOV) of the camera was determined to be 484 mm across all 1936 spatial pixels, resulting in a spatial resolution of approximately 0.25 mm/pixel along the direction of the scanning line. For the moving direction of the imaging unit along a predetermined distance, the spatial resolution depends on the translation speed of the stage and the number of total scans. For example, using a moving speed of 20 mm/s, it will require 40 s to scan 800 lines for one-way travel over an 800-mm distance, resulting in an approximately 1 mm/pixel spatial resolution. To synchronize continuous line-scan image acquisition and translation stage movement, the moving speed of the stage is automatically determined by the system software based on the independently selected exposure time of the camera (i.e., low speed for long exposure time and high speed for short exposure time). Based on test results with the current laptop using selected exposure times and corresponding stage moving speeds, an empirical reciprocal relationship was found between the moving speed (V in mm/s) and the exposure time (T in s) (i.e., V=1/T). For example, for the exposure times of 0.05 and 0.1 s, the moving speeds were determined to be 20 and 10 mm/s, respectively. Note that this relationship may be affected by actual frame rate and USB data transfer speed when the camera is connected to different computers.

2.2 Hyperspectral image processing

This section describes procedures for processing raw hyperspectral reflectance and fluorescence images, including spectral and spatial averaging, flat-field correction, background removal, and image registration. Dragoon lettuce (*Lactuca sativa*) samples planted in 10 cm square soil pots were used to demonstrate the image processing results. Note that all images presented in this paper were acquired using the continuous scanning mode. The hyperspectral reflectance and fluorescence images were acquired from the standard panel and the lettuce in the pot using 0.05 s exposure time, 1200 spatial pixels, and 350 line scans over a 350 mm distance, which generated two 1200×350×300 (X×Y×λ) raw hypercubes. Initial smoothing of the spectral and spatial data was performed on the 3-D raw images by averaging across groups of three neighboring pixels in the spectral dimension (λ) and also across groups of four neighboring pixels in the camera's scanning line direction (X), which created two 300×350×100 reduced hypercubes with a spatial resolution of 1 mm/pixel in both X and Y dimensions and a spectral interval of 5.94 nm. Using the VNIR reflectance values of the standard reference panel (i.e., 100%), flat-field correction (Kim et al., 2001) was conducted to convert the intensity values in the averaged image to relative reflectance values (0–100%).

To remove the background of the lettuce, correlation analysis was used to identify an optimal two-band reflectance ratio (i.e., Rλ1/Rλ2, where Rλn denotes single-band reflectance image at

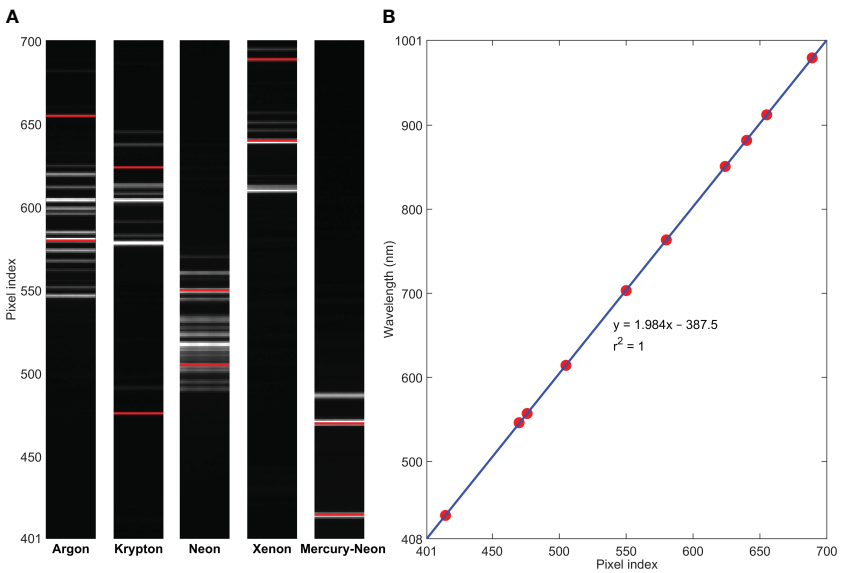


FIGURE 3
Spectral calibration for the hyperspectral camera using **(A)** standard pencil calibration lamps (two spectral lines selected from each lamp) and **(B)** a linear regression model.

wavelength of λn) to segregate the leaves from the soil. Reflectance spectra of the lettuce and the soil were extracted from regions of interest (ROIs) that were manually selected from a single-band image at 815 nm (high reflectance for lettuce). The spectra of the lettuce and the soil were labeled with 1 and 2, respectively. Correlation coefficients were calculated between all two-band ratios of the ROI spectra and the label values. The ratio image that gave the highest correlation was converted to a binary mask image by a single threshold value computed using Otsu's method (Otsu, 1979). To align the reflectance and fluorescence images of the same sample, a white paper printed with a grid of black dots was used as a calibration template for feature-based image registration with reflectance as a fixed image and fluorescence as a moving image. In-house programs developed by MATLAB were used to execute all the image processing procedures.

2.3 Application to plant drought stress detection

The hyperspectral plant health monitoring system is intended to be used for early detection of abiotic stresses (e.g., drought, overwatering, and nutrient deficiencies) and plant diseases (e.g., bacterial, fungal, and viral) for crops grown under controlled environments. A pilot study for detecting plant drought stress was conducted to demonstrate the performance and capability of the system, using Dragoon lettuce, a mini green romaine previously grown on the International Space Station. Each individual lettuce plant, of twelve total plants divided into two sets of six, was grown in mix of soil and arcillite (v:v=3:7) in a 10 cm square pot for a total of 28 days after seeding. The plants were cultivated under an air temperature of 23°C, a relative humidity of 65%, a CO_2 concentration of 3000 ppm (simulating spacecraft cabin air), and a photoperiod of 16 h light and 8 h dark using a LED plant grow light providing a photosynthetic photon flux density (PPFD) of 300 μmol m^{-2} s^{-1} and a spectral output of 90% red, 1% green, and 9% blue. Note that in this experiment, the plant samples were grown under a LED grow light that was not in the imaging chamber. Two different moisture treatments for the soil were started on Day 12 after planting (i.e., stress Day 1) so that six lettuce samples were grown under well-watered conditions (control) and six were under-watered (drought stress). Each set of six pots was arranged in a 2×3 array on a tray placed inside the growth chamber. Moisture in each tray was controlled by an automated watering system that periodically watered the plants with nutrient solution to maintain a pre-determined volumetric moisture content (VMC), which was measured by a soil moisture sensor. The VMCs for the control and drought treatments were set as 50% and 30%, respectively. Thus, the average moisture content of the drought pots contained 100 ml less water than the controls. After introduction of the drought stress, hyperspectral images were taken over 13 days within a period of three weeks (i.e., Week 1: Days 1–4, Week 2: Days 7–11, and Week 3: Days 14–17). Reflectance and fluorescence images were collected from each tray using 1600 spatial pixels and 550 line scans across a 550 mm distance under a common camera exposure time of 0.05 s. It took approximately 90 s to obtain two 1600×550×300 raw hypercubes. Image processing was performed using the procedures described in Section 2.2, which generated two 300×360×100 reduced hypercubes for each set of six lettuce samples.

Three methods, including leaf area, band-ratio, and machine learning, were used for the drought stress detection. For all 13 sampling days in the three-week period, total leaf areas for each set of six samples were estimated daily by counting the pixel numbers in the plant mask images. Correlation analysis was then used to identify a two-band ratio for the drought stress detection. Reflectance spectra of control and drought lettuce samples were extracted from masked images for each day, which were then grouped into individual weeks for the correlation analysis. Last, the potential of machine learning method for early detection of drought stress was investigated using reflectance spectra of lettuce. A pixel average-window method was first used to remove the plant areas with large variations (e.g., leaf margins) and reduce the number of spectra used in machine learning classifications. All the lettuce pixels in the masked R815 images (high reflectance for lettuce) were grouped into 3×3 pixel windows, in which mean (M) and standard deviation (SD) of the pixel intensities were calculated. In each window, if there were more than 10% of nine pixels (i.e., one or more pixels) with the reflectance intensities beyond the range of M ± 3SD, the whole window was removed for further analysis. The nine spectra extracted from each remaining window were averaged in the spatial domain. All mean spectra labeled with control and drought were used for the machine learning classifications. Each of nine labeled datasets from stress Days 1 to 11 was input to the Classification Learner app in MATLAB, in which seven optimizable classifiers, including Naive Bayes, decision tree, ensemble, k-nearest neighbor (KNN), support vector machine (SVM), neural network (NN), and discriminant analysis, were used to compare the classification accuracies. Hyperparameter optimization functions within the app were used for automated selection of the hyperparameters for all the models to minimize the classification error. To simplify the evaluation of misclassification costs and model training and validation, equal penalty was assigned to all misclassifications. Accuracies of the seven classification models for the dataset on each stress day in the first two weeks were evaluated using a five-fold cross-validation method. To minimize variations from random dataset partitioning, training and validation of each model was repeated for ten times. The average cross-validation accuracy over the ten runs was used as the overall accuracy of each model.

3 Results and discussion

3.1 Hyperspectral image processing results

Figure 4 shows reflectance image processing results for a lettuce sample. After spectral and spatial averaging, a reduced hypercube (Figure 4B) was first generated from the raw hypercube (Figure 4A). Then, the reduced hypercube was converted to a reflectance hypercube (Figure 4C) *via* the flat-field correction. The averaged raw and reflectance spectra of the panel, the lettuce, and the soil at three selected locations are plotted in Figures 4D, E, respectively.

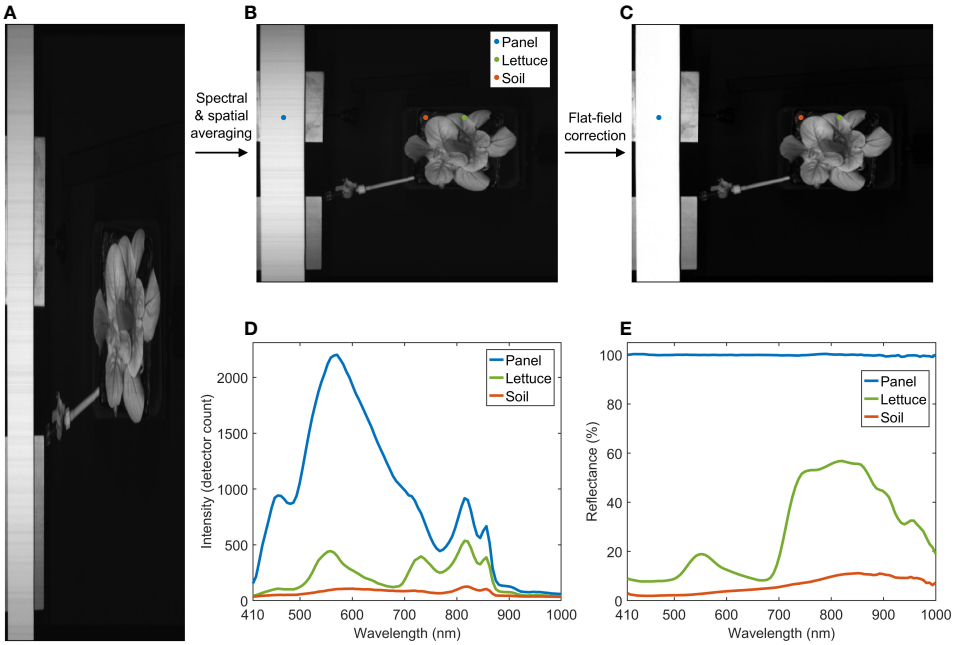


FIGURE 4
Spectral and spatial averaging and flat-field correction for hyperspectral reflectance image of a lettuce sample. **(A)** raw, **(B)** reduced, and **(C)** reflectance images at 815 nm, and **(D)** raw and **(E)** reflectance spectra of panel, lettuce, and soil.

The lettuce sample shows a typical vegetation reflectance spectrum in the VNIR region. The reflectance signals of the lettuce beyond 850 nm tend to drop and fluctuate toward 1000 nm due to relatively low output of the 915 nm LEDs, spectrograph-produced second-order effect (Kim et al., 2011), and a low quantum efficiency of the detector in this wavelength range.

Removing the background of a lettuce sample in a soil pot is demonstrated in Figure 5. The hyperspectral reflectance image was collected and processed in the same way as those shown in Figure 4, but with the addition of image cropping to center and contain the whole plant pot in a reduced hypercube size of 130×130×100. The ratio between 559 and 678 nm (i.e., R559/R678) gave the maximum absolute correlation coefficient of −0.91 (Figure 5A). The selected two wavelengths are marked on top of the mean spectra of the lettuce and the soil (Figure 5B). Single-band images at 559 and 678 nm are shown in Figures 5C, D, respectively. The two mean spectra exhibit opposite trends between 559 and 678 nm, indicating that these wavelengths are effective selections for the correlation analysis and band-ratio method for background removal. As shown in Figure 5E, the contrast between the leaves and the background is greatly enhanced in the ratio image of R559/R678. Lastly, a mask image (Figure 5F) was obtained by applying a single threshold value to the ratio image. Note that although the optimal wavelength pair was selected based specifically on the reflectance spectra of the lettuce and the soil, the ratio of R559/R678 is also effective for removing other objects in the background, such as the pot, tray, and drip irrigation tubing.

Figure 6 illustrates image registration, masking, and spectral extraction for the hyperspectral reflectance and fluorescence images collected from the same lettuce sample shown in Figure 5. An overlap of R815 and F732 (i.e., single-band fluorescence image at 732 nm) images at their original spatial positions are shown in Figure 6A, which clearly illustrates misalignment of the reflectance and fluorescence images of the same scene. The registration result from the calibration template showed that a simple horizontal translation (10-pixel shift to the left for this case) is adequate to align the fluorescence image with the reflectance image. Thus, the double-image effect disappeared in the overlap of the registered R815 and F732 images (Figure 6B). Note that same camera exposure time of 0.05 s (thus same stage moving speed in continuous scanning mode) was used to acquire both reflectance and fluorescence images in this example. If different exposure times are used, then step-by-step scanning mode or more advanced image registration techniques may be needed to align the images. After the registration, the mask image created using R559/R678 (Figure 6C) can be used to mask both reflectance and fluorescence images (Figures 6D, E, G, H). In addition, a pseudo RGB image (registered and masked shown in Figures 6F, I, respectively) was generated using red, green, and blue bands of the reflectance hypercube to provide natural color and appearance for the lettuce sample. Based on the masked images, spectra can be extracted from all the lettuce pixels for further analysis. Mean and standard deviation (SD) reflectance and fluorescence spectra of the lettuce sample are plotted in Figures 6J, K, respectively. Due to chlorophyll a in the lettuce leaf tissue, the reflectance spectra show an absorption peak at 672 nm and the fluorescence spectra show two red emission peaks at 690 and 732 nm. In addition, owing to phenolic compounds in the leaves (Chappelle et al., 1991), two broad blue and green emission peaks with low fluorescence intensities were observed around 450 and 530 nm, respectively.

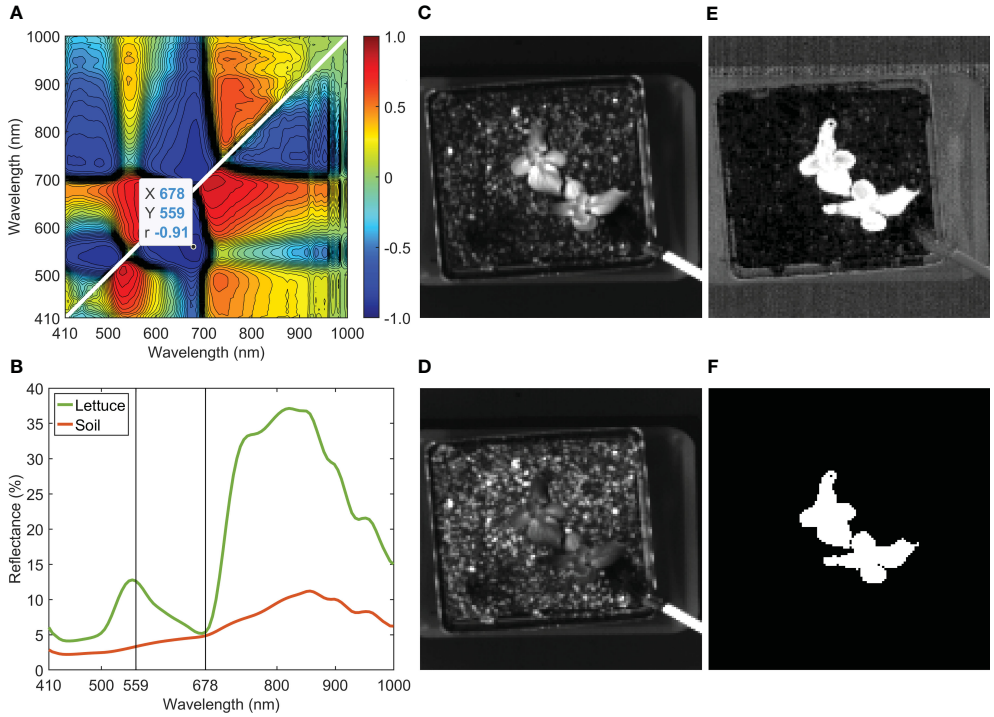


FIGURE 5
Background removal of a lettuce sample: **(A)** correlation analysis for band-ratio selection to segregate lettuce from soil, **(B)** mean reflectance spectra of lettuce and soil, reflectance images at **(C)** 559 nm and **(D)** 678 nm, **(E)** ratio image (R559/R678), and **(F)** mask image generated using R559/R678.

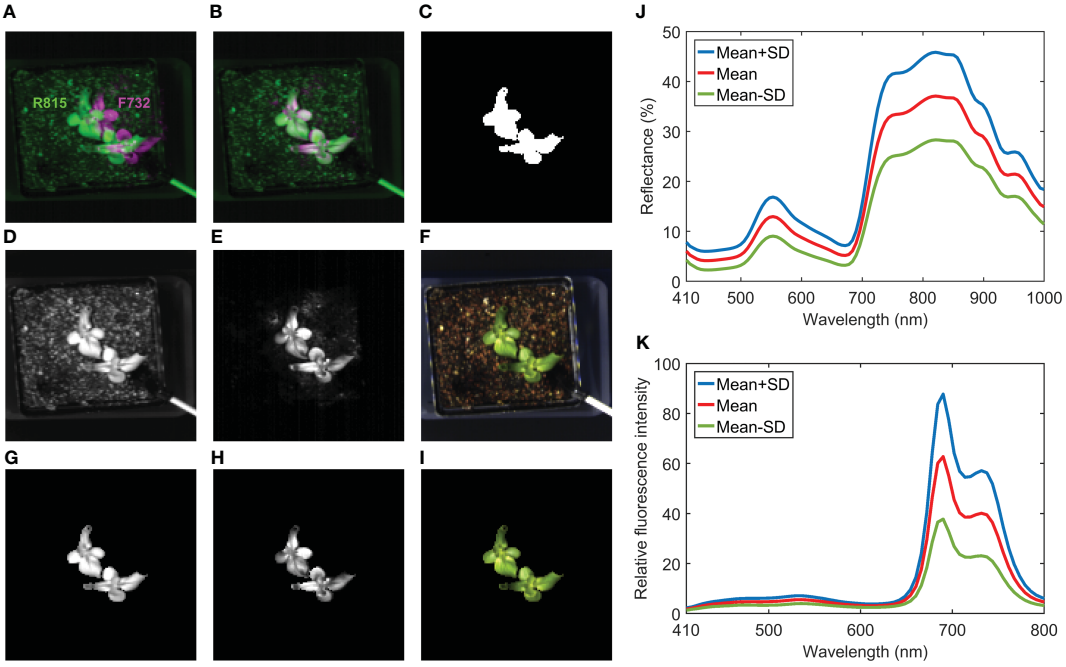


FIGURE 6
Image registration for hyperspectral reflectance and fluorescence images of a lettuce sample: overlap of **(A)** original and **(B)** registered R815 and F732 images, **(C)** mask image generated using R559/R678, **(D)** registered and **(G)** masked R815 images, **(E)** registered and **(H)** masked F732 images, **(F)** registered and **(I)** masked pseudo RGB images, and **(J)** reflectance and **(K)** fluorescence spectra extracted from masked images.

3.2 Plant drought stress detection results

3.2.1 Leaf area results

Results for the leaf area method are plotted in Figure 7. The average pixel areas between control and drought plants on stress Days 3, 4, 7 and 8 were compared using a two-sample t-test. The pixel areas from drought plants on stress Days 3 and 4 were not significantly different from the controls. However, they were significantly different on stress Day 7 (13% less leaf area, $p<0.02$) and Day 8 (29% less leaf area, $p<0.001$). The decrease in leaf area caused by drought persisted until harvest.

3.2.2 Band-ratio results

Correlation coefficients calculated using the reflectance spectra in Week 1 were low (Figure 8A), with a maximum of 0.22 at R821/R869. When Week 2 data were used, the maximum correlation increased to 0.55 at R690/R702 (Figure 8B). For each day in Week 2, the contour plot of correlation coefficients showed a similar pattern to that from using the data of the whole week. Mean reflectance spectra of the control and drought samples in Weeks 1 and 2 were normalized at 702 nm and are plotted across the wavelength range of 540–740 nm in Figures 8C, D, respectively. In Week 1, there was no notable spectral difference between 690 and 702 nm for the two moisture treatments, which is the reason for the low correlation coefficient (i.e., −0.07) at R690/R702 (Figure 8A). In Week 2, however, the mean spectrum of the drought samples showed higher reflectance at 690 nm than that of the control samples. Both selected wavelengths of 690 and 702 nm are in the red edge spectral region (690–740 nm), in which reflectance of the leaves is sensitive to the change of the chlorophyll content in green plants (Lowe et al., 2017). Single-band (R690 and R702) and band-ratio (R690/R702) images of a control and a drought lettuce samples on stress Days 4 and 9 are shown in Figures 8C, D, respectively. On Day 4, the two ratio images exhibited similar intensity patterns for the control and drought samples. On Day 9, however, the ratio image of the drought sample showed higher intensities over most of the leaf areas, in great contrast to the lower intensities of the control sample leaf areas.

Figure 9 shows pseudo RGB and band-ratio images of all the lettuce samples in the first two weeks. The pseudo RGB images show that there were no apparent differences for the natural color and appearance of the lettuce grown under the two moisture treatments. For the ratio images from stress Days 1 to 4 in Week 1, no obvious intensity differences were observed between the control and drought samples. The first notable difference for the drought samples appeared in the ratio image on stress Day 7, in which all six plants showed higher ratio intensities than the controls. In Week 2, the ratio values of R690/R702 of the drought samples tended to increase over the whole leaf areas from stress Days 7 to 11, while those of the controls generally remained unchanged, except for some small leaf areas. From stress Days 9 to 11, the ratio images of the three drought samples on the bottom row showed lower intensities than those of the three on the top row. Such variations can probably be attributed to the non-uniform water supply received by each individual pot. As the drought samples were grown into stress Week 3, the ratio intensity variations of the six plants increased and their ratio intensity differences with the six control samples decreased (results not shown). These results suggest that the band-ratio method can detect the drought stress at approximately the same time as the leaf area method, which may not be adequate for the goal of early stress detection.

3.2.3 Machine learning results

Accuracies for differentiating the control and drought lettuce samples using seven optimized classification models are shown in Figure 10D using the five-fold cross-validation results from stress Days 4 (Figure 10A) and 7 (Figure 10E). For both days, the Naive Bayes and the discriminant classifiers generated the lowest (worse than 65%) and the highest (better than 90%) accuracies, respectively, and the accuracies of other five classifiers fell between those two. Meanwhile, the accuracies of all seven classifiers using the reflectance spectra on

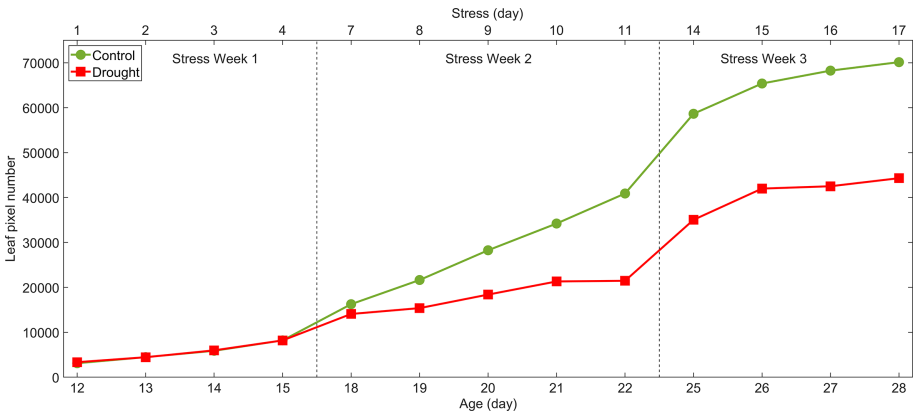


FIGURE 7
Leaf area method for detection of drought stress of lettuce: total leaf areas on each sampling day estimated by pixel numbers in mask images of control and drought lettuce samples.

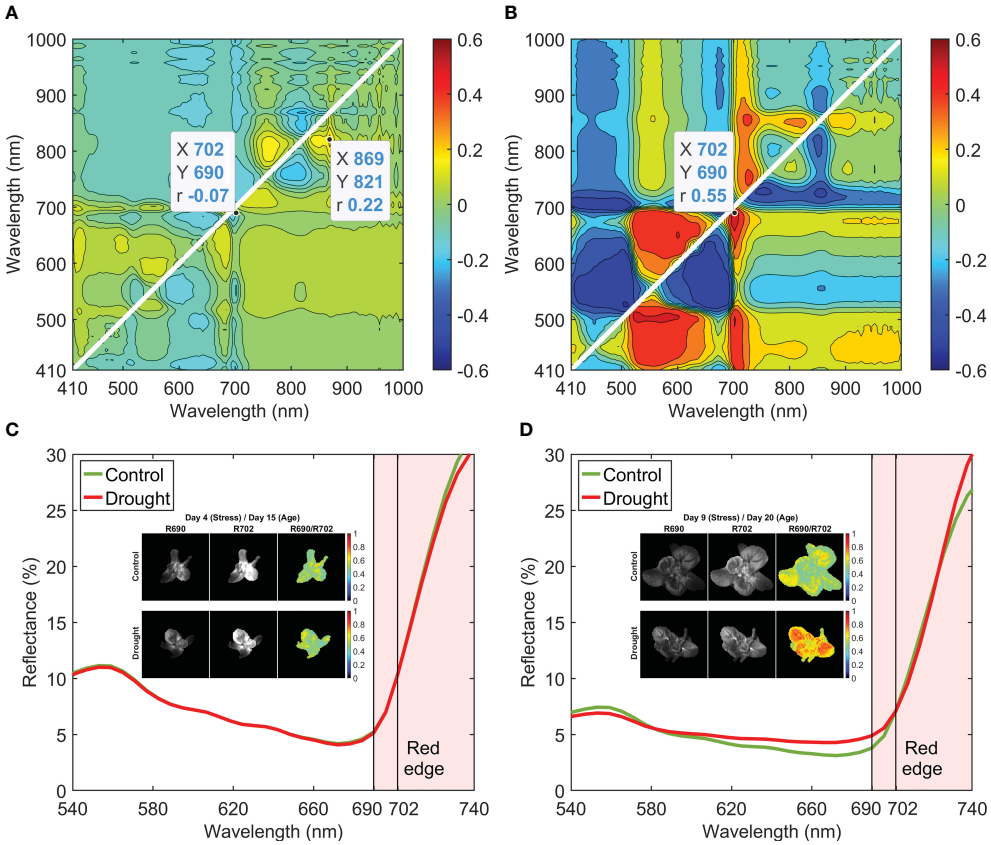


FIGURE 8
Correlation analysis for band-ratio selection to differentiate control and drought lettuce samples using reflectance spectra from **(A)** Week 1 (stress Days 1–4) and **(B)** Week 2 (stress Days 7–11). Normalized mean spectra and example single-band and band-ratio images for Weeks 1 and 2 are shown in **(C, D)**, respectively.

Day 7 are consistently higher than those using the data on Day 4. Confusion matrices and receiver operating characteristic (ROC) curves on these two selected days using the discriminant classifier are shown in Figures 10B, F and Figures 10C, G, respectively. The classification accuracies for Days 4 and 7 are 94.3% and 98.5%, respectively, and the areas under both ROC curves are better than 0.98. Similar results were obtained for other days in the first two weeks. Since the discriminant classifier gave the best overall classification performance, it was selected for differentiating the control and drought lettuce samples in all nine test days in the first two weeks, and the results are summarized in Figure 10H. In Week 1, the classification accuracies gradually increased from Day 1 (90.7%) to Day 4 (94.3%). In Week 2, all the accuracies

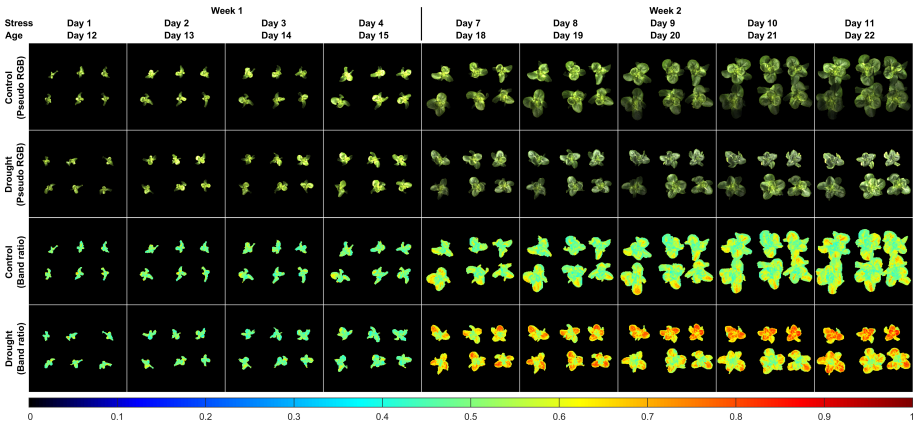


FIGURE 9
Reflectance band-ratio method for detection of drought stress of lettuce: pseudo RGB and ratio (R690/R702) images of samples grown under well-watered (control) and under-watered (drought) conditions in the first two sampling weeks.

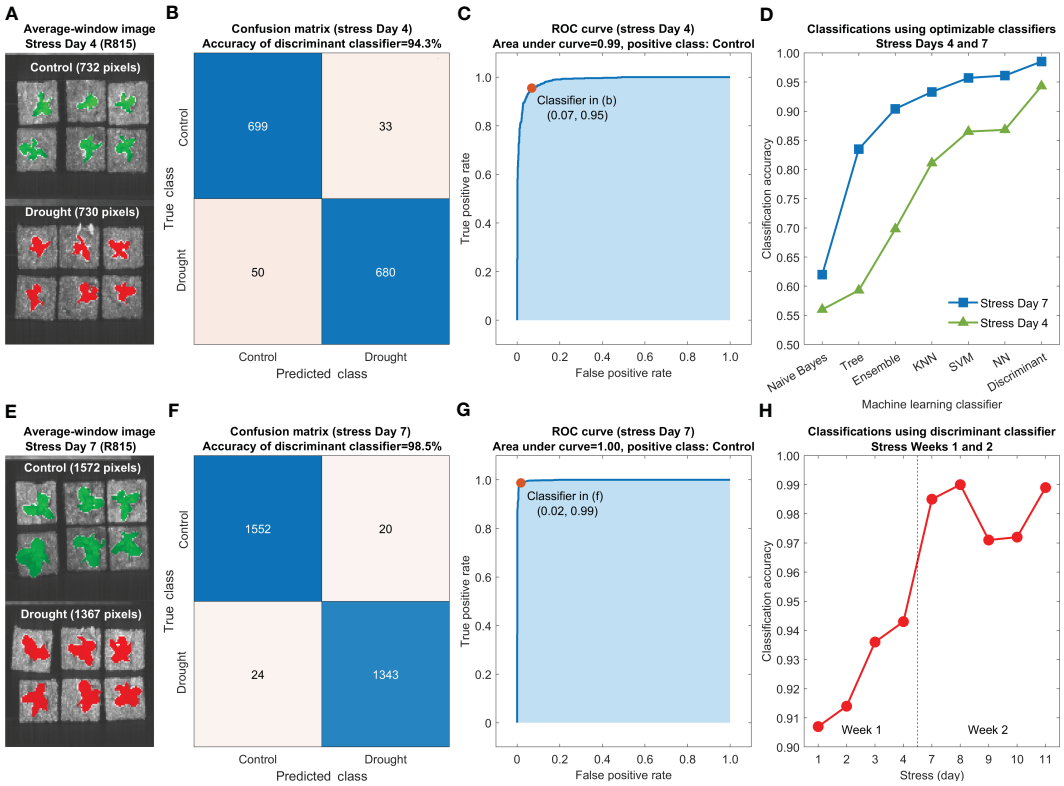


FIGURE 10
Machine learning classification method for early detection of drought stress of lettuce using reflectance spectra extracted from hyperspectral images of the leaves: average-window R815 images on stress Days **(A)** 4 and **(E)** 7, confusion matrices for stress Days **(B)** 4 and **(F)** 7 and ROC curves for stress Days **(C)** 4 and **(G)** 7 using discriminant classifiers, **(D)** classification accuracies using seven optimized classifiers for stress Days 4 and 7, and **(H)** classification accuracies using discriminant classifiers for all nine stress days in the first two weeks.

were higher than 97.0% with a maximum of 99.0% on Day 8. Results from this preliminary experiment suggest that the machine learning method using an optimized discriminant classifier based on VNIR hyperspectral reflectance images is promising for early detection of drought stress for lettuce without any visible symptoms and leaf size differences.

The results above demonstrate that the hyperspectral reflectance imaging based on the broadband LED light can be potentially used for plant health monitoring. The VNIR LED light provides an alternative to the halogen light commonly used for the reflectance measurement, which can avoid excessive heat from the halogen light projected to the tender plants under a close-range imaging setup. As the next step of this collaborative project, full-scale plant experiments with replications will be conducted to collect image data from salad crops of multiple species (e.g., lettuce, pak choi, mizuna, and radish) grown under different abiotic and biotic stress treatments. Both separate and combined use of the reflectance and fluorescence data for assessing plant vigor will be investigated and compared. Spectral and image fusion algorithms will be developed toward the goal of early detection of plant stresses and diseases. The hyperspectral plant health monitoring system has great potential to provide timely and useful information for optimization of the growth environment and early mitigation of plant stress and disease in space crop production systems, as well as for other applications in controlled-environment agriculture.

4 Conclusions

As a first fruit of a collaborative project between USDA ARS and NASA KSC, a compact hyperspectral imaging prototype system was developed and preliminarily tested for monitoring plant health in controlled-environment space crop production. The prototype system can acquire both hyperspectral reflectance and fluorescence images in the visible and near-infrared region within a single imaging cycle, which can provide rich spectral and spatial information to potentially carry out early detection of abiotic stresses and plant diseases in pick-and-eat salad crops. Compact and lightweight hardware components, including two LED line lights, a hyperspectral camera, and a motorized stage, were used to build the imaging unit to ensure it can be integrated into the confined space of a growth chamber to conduct overhead sensor-to-sample imaging. The broadband and UV-A LED lights project a narrow and structured beam to illuminate plant samples during rapid image acquisition. Use of VNIR LED light instead of traditional halogen light for reflectance can avoid excessive heat projected to the plant samples under a close-range imaging setup. The in-house developed system control software provides a user-friendly interface for plant scientists to conduct imaging experiments. The performance and capability of the developed system was demonstrated in a pilot study on plant drought stress detection for Dragoon lettuce. A reflectance

band-ratio method based on two wavelengths selected in the red edge spectral region was found to be inadequate for early stress detection, as it could only differentiate control and drought samples at the same stage of growth as was possible from traditional leaf area estimation. A machine learning method using an optimized discriminant classifier based on VNIR reflectance spectra showed promise for early detection of drought stress on lettuce leaves lacking visible symptoms and size differences. To fully utilize the potential of hyperspectral reflectance and fluorescence imaging techniques to achieve the goal of detecting early onset of plant stresses and diseases in the space crop production, full-scale experiments on multiple species and treatments and the development of more advanced spectral and image analysis and fusion algorithms are planned as the next step of this project.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

JQ: conceptualization, methodology, investigation, formal analysis, writing—original draft, writing—review and editing. OM: data curation, formal analysis, investigation, writing—review and editing. MN: validation, data curation. JF: validation, data curation. AO: data curation, investigation, writing—review and editing. KW: project administration, supervision, funding acquisition. RF: project administration, supervision, funding acquisition. IB: methodology, validation, investigation. DC: resources, writing—review and editing. MK: conceptualization, methodology, investigation, writing—review and editing, project administration, supervision, funding acquisition. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bai, G., Ge, Y., Scoby, D., Leavitt, B., Stoerger, V., Kirchgessner, N., et al. (2019). NU-spidercam: a large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing, and agronomic research. *Comput. Electron. Agric.* 160, 71–81. doi: 10.1016/j.compag.2019.03.009

Bao, Y., Shah, D. S., and Tang, L. (2018). 3D perception-based collision-free robotic leaf probing for automated indoor plant phenotyping. *Trans. ASABE* 61, 859–872. doi: 10.13031/trans.12653

Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., et al. (2018). Specim IQ: evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. *Sensors* 18, 441. doi: 10.3390/s18020441

Behmann, J., Mahlein, A., Paulus, S., Dupuis, J., Kuhlmann, H., Oerke, E., et al. (2016). Generation and application of hyperspectral 3D plant models: methods and challenges. *Mach. Vision Appl.* 27, 611–624. doi: 10.1007/s00138-015-0716-8

Chappelle, E. W., McMurtrey, J. E., and Kim, M. S. (1991). Identification of the pigment responsible for the blue fluorescence band in the laser induced fluorescence (LIF) spectra of green plants, and the potential use of this band in remotely estimating rates of photosynthesis. *Remote Sens. Environ.* 36 (3), 213–218. doi: 10.1016/0034-4257(91)90058-E

Ge, Y., Bai, G., Stoerger, V., and Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Comput. Electron. Agric.* 127, 625–632. doi: 10.1016/j.compag.2016.07.028

Huang, P., Luo, X., Jin, J., Wang, L., Zhang, L., Liu, J., et al. (2018). Improving high-throughput phenotyping using fusion of close-range hyperspectral camera and low-cost depth sensor. *Sensors* 18, 2711. doi: 10.3390/s18082711

Jiang, Y., Li, C., Robertson, J. S., Sun, S., Xu, R., and Paterson, A. H. (2018). GPhenoVision: a ground mobile system with multi-modal imaging for field-based high throughput phenotyping of cotton. *Sci. Rep.* 8, 1213. doi: 10.1038/s41598-018-19142-2

Kim, M. S., Chao, K., Chan, D. E., Jun, W., Lefcout, A. M., Delwiche, S. R., et al. (2011). Line-scan hyperspectral imaging platform for agro-food safety and quality evaluation: system enhancement and characterization. *Trans. ASABE* 54, 703–711. doi: 10.13031/2013.36473

Kim, M. S., Chen, Y., and Mehl, P. M. (2001). Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Trans. ASAE* 44, 721–729. doi: 10.13031/2013.6099

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111. doi: 10.3390/s141120078

Lien, M. R., Barker, R. J., Ye, Z., Westphall, M. H., Gao, R., Singh, A., et al. (2019). A low-cost and open-source platform for automated imaging. *Plant Methods* 15, 6. doi: 10.1186/s13007-019-0392-1

Lowe, A., Harrison, N., and French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* 13, 80. doi: 10.1186/s13007-017-0233-z

Massa, G. D., Wheeler, R. M., Morrow, R. C., and Levine, H. G. (2016). Growth chambers on the international space station for large plants. *Acta Hortic.* 1134, 215–222. doi: 10.17660/ActaHortic.2016.1134.29

Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., and Scheunders, P. (2017). Close range hyperspectral imaging of plants: a review. *Biosyst. Eng.* 164, 49–67. doi: 10.1016/j.biosystemseng.2017.09.009

Monje, O., Nugent, M. R., Hummerick, M. E., Dreschel, T. W., Spencer, L. E., Romeyn, M. W., et al. (2019). *New frontiers in food production beyond LEO. 49th* (international conference on environmental systems, Boston, MA, USA), ICES–2019-260. Available at: https://hdl.handle.net/2346/84471.

Nguyen, C., Sagan, V., Bhadra, S., and Moose, S. (2023). UAV multisensory data fusion and multi-task deep learning for high-throughput maize phenotyping. *Sensors* 23, 1827. doi: 10.3390/s23041827

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Systems Man Cybernetics* 9 (1), 62–66. doi: 10.1109/TSMC.1979.4310076

Pérez-Bueno, M. L., Pineda, M., Cabeza, F. M., and Barón, M. (2016). Multicolor fluorescence imaging as a candidate for disease detection in plant phenotyping. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01790

Qin, J., Kim, M. S., Chao, K., Chan, D. E., Delwiche, S. R., and Cho, B. (2017). Line-scan hyperspectral imaging techniques for food safety and quality applications. *Appl. Sci.* 7 (2), 125. doi: 10.3390/app7020125

Rousseau, C., Belin, E., Bove, E., Rousseau, D., Fabre, F., Berruyer, R., et al. (2013). High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis. *Plant Methods* 9, 17. doi: 10.1186/1746-4811-9-17

Stergar, J., Hren, R., and Milanič, M. (2022). Design and validation of a custom-made laboratory hyperspectral imaging system for biomedical applications using a broadband led light source. *Sensors* 22, 6274. doi: 10.3390/s22166274

Unger, P., Sekhon, A. S., Chen, X., and Michael, M. (2022). Developing an affordable hyperspectral imaging system for rapid identification of escherichia coli O157:H7 and listeria monocytogenes in dairy products. *Food Sci. Nutr.* 10 (4), 1175–1183. doi: 10.1002/fsn3.2749

Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., and Hawkesford, M. J. (2017). Field scanalyzer: an automated robotic field phenotyping platform for detailed crop monitoring. *Funct. Plant Biol.* 44, 143–153. doi: 10.1071/fp16163

Wang, L., Jin, J., Song, Z., Wang, J., Zhang, L., Rehman, T. U., et al. (2020). LeafSpec: an accurate and portable hyperspectral corn leaf imager. *Comput. Electron. Agric.* 169, 105209. doi: 10.1016/j.compag.2019.105209

White, J. W., and Conley, M. M. (2013). A flexible, low-cost cart for proximal sensing. *Crop Sci.* 53, 1646–1649. doi: 10.2135/cropsci2013.01.0054

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01111

Zeidler, C., Zabel, P., Vrakking, V., Dorn, M., Bamsey, M., Schubert, D., et al. (2019). The plant health monitoring system of the EDEN ISS space greenhouse in Antarctica during the 2018 experiment phase. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01457

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop phenomics: current status and perspectives. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.0071

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 July 2023
DOI 10.3389/fpls.2023.1188286

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Sijia Yu,
Rutgers, The State University of New Jersey - Busch Campus, United States
Ronnie Concepcion II,
De La Salle University, Philippines

*CORRESPONDENCE
Shengyong Xu
xsy@mail.hzau.edu.cn

RECEIVED 17 March 2023
ACCEPTED 30 June 2023
PUBLISHED 14 July 2023

CITATION
Qiao Y, Liao Q, Zhang M, Han B, Peng C, Huang Z, Wang S, Zhou G and Xu S (2023) Point clouds segmentation of rapeseed siliques based on sparse-dense point clouds mapping.
Front. Plant Sci. 14:1188286.
doi: 10.3389/fpls.2023.1188286

COPYRIGHT
© 2023 Qiao, Liao, Zhang, Han, Peng, Huang, Wang, Zhou and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Point clouds segmentation of rapeseed siliques based on sparse-dense point clouds mapping

Yuhui Qiao[1,2], Qingxi Liao[1,2], Moran Zhang[1,2], Binbin Han[3], Chengli Peng[4], Zhenhao Huang[1,2], Shaodong Wang[1,2], Guangsheng Zhou[5] and Shengyong Xu[1,2]*

[1]College of Engineering, Huazhong Agricultural University, Wuhan, China, [2]Key Laboratory of Agricultural Equipment for the Middle and Lower Reaches of the Yangtze River, Ministry of Agriculture, Huazhong Agricultural University, Wuhan, China, [3]School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China, [4]School of Geosciences and Info-Physics, Central South University, Changsha, China, [5]College of Plant Science & Technology, Huazhong Agricultural University, Wuhan, China

In this study, we propose a high-throughput and low-cost automatic detection method based on deep learning to replace the inefficient manual counting of rapeseed siliques. First, a video is captured with a smartphone around the rapeseed plants in the silique stage. Feature point detection and matching based on SIFT operators are applied to the extracted video frames, and sparse point clouds are recovered using epipolar geometry and triangulation principles. The depth map is obtained by calculating the disparity of the matched images, and the dense point cloud is fused. The plant model of the whole rapeseed plant in the silique stage is reconstructed based on the structure-from-motion (SfM) algorithm, and the background is removed by using the passthrough filter. The downsampled 3D point cloud data is processed by the DGCNN network, and the point cloud is divided into two categories: sparse rapeseed canopy siliques and rapeseed stems. The sparse canopy siliques are then segmented from the original whole rapeseed siliques point cloud using the sparse-dense point cloud mapping method, which can effectively save running time and improve efficiency. Finally, Euclidean clustering segmentation is performed on the rapeseed canopy siliques, and the RANSAC algorithm is used to perform line segmentation on the connected siliques after clustering, obtaining the three-dimensional spatial position of each silique and counting the number of siliques. The proposed method was applied to identify 1457 siliques from 12 rapeseed plants, and the experimental results showed a recognition accuracy greater than 97.80%. The proposed method achieved good results in rapeseed silique recognition and provided a useful example for the application of deep learning networks in dense 3D point cloud segmentation.

KEYWORDS

rapeseed siliques, 3D Reconstruction, sparse-dense point clouds mapping, point clouds segmentation, silique recognition

1 Introduction

Rapeseed is one of the main oil crops in the world (Shyam et al., 2012). The yield of rapeseed is mainly determined by the number of siliques per plant, the number of seeds per silique, and the thousand seed weight (Tang et al., 2020). Among them, the number of siliques per plant has the highest correlation with yield. However, currently, the counting of rapeseed siliques is still done manually, which is not only time-consuming and laborious but also usually has a high cost. This to some extent has limited the rapid progress of rapeseed-related research. Therefore, the intelligent recognition, segmentation, and counting of rapeseed siliques have important research significance.

The 2D image-based silique counting method is fast in detection and high in throughput (Du et al., 2023). For example, Liu et al. proposed an improved K-means-based method for counting wheat ears. Utilizing the clustering of color features, the number of sub-regions within the clustered region was employed as a mean of approximating the wheat crop count. The accuracy of wheat ear counting reached 94.69% (Liu et al., 2019). Dandrifosse, S proposed a deep learning method. It is an unsupervised learning method using the YOLOv5 model on and the cutting edge DeepMAC segmentation method capable of counting and segmenting wheat ears (Dandrifosse et al., 2022). Wang, proposed an improved efficient entdet-d0 model. The introduction of the convolutional block attention module (CBAM) in the model allows the model to refine the features, focus more on the wheat ears and suppress other useless background information. The problem of overlap in wheat ears images can be solved in wheat ears detection and counting (Wang et al., 2021). Zhang proposed a high-precision wheat head detection model with strong generality based on a single-stage network structure. The number of wheats can be detected quickly (Zhang et al., 2022). Liu et al. accurately calculated the number of canola siliques by scatter treatment and image analysis of piled-up siliques(Ren-feng et al., 2020). However, there are high requirements for automated loading and unloading devices. Zhao proposed a P2PNet-Soy method that maximizes the performance of the model in soybean counting and localization by adjusting the architecture and subsequent post-processing. It achieves higher accuracy than the original P2PNet (Zhao et al., 2023). However, traditional methods for 2D image processing have difficulty in solving the occlusion problem due to the lack of high-level semantic features. The existing counting methods based on deep learning of 2D images are also ineffective in solving the occlusion problem. In contrast, precise 3D point clouds data could obtain from various sensors such as Time-of-Flight (ToF), structured light cameras, and 3D laser scanners, which can provide more detailed spatial information about plants to solve the above problems.

High resolution and accurate 3D point cloud data allow for more accurate 3D morphological parameters of plants (Ni et al., 2021). 3D reconstruction is the basis for 3D three-dimensional phenotyping studies of crops, but the requirements for algorithms are high. For instance, with the Kinect V2 (a kind of camera), Xu et al. captured images of rapeseed branches from four different perspectives to reconstruct the rapeseed branches. Then they used the super-green segmentation algorithm to extract the siliques, removed the largest connected domain using opening operation, and identified and located individual siliques through Euclidean clustering (Shengyong et al., 2019). However, due to the swinging effect caused by the tall and soft rapeseed plants, the measurement errors of ToF and structured light sensors used widely range from millimeter to centimeter level (Tao and Zhou, 2017). Moreover, the slender shape of rapeseed siliques requires high spatial resolution for depiction. Therefore, the 3D data of rapeseed plants obtained by ToF or structured light sensors may be less accurate and reliable. 3D scanners (e.g., laser and radar) will give high-precision point clouds. There are more than 100,000 data points for an entire plant. Each plant organ has more than 100,000 to 30,000 data points. Although the point clouds obtained by this method are of high quality, they suffer from problems such as expensive equipment (Jin et al., 2021). A low-cost and high-precision method to construct 3D point clouds of plants can be achieved using smartphone and SFM algorithm (Marzulli et al., 2020). However, the cumbersome process of taking photos from multiple angles greatly affects the efficiency of acquiring point clouds. Therefore, this study proposes a video-based SFM 3D reconstruction method for an entire mature rapeseed plant, aiming at achieving low-cost, high-accuracy, and high-efficiency 3D reconstruction. This makes it possible to divide and count whole rapeseed siliques with high precision. However, the numerous siliques of the mature rapeseed are scattered and partially overlapping, making it difficult to identify them accurately. Therefore, high-precision identification and segmentation of rapeseed siliques remain a challenge (Li et al., 2022). Although the rapid development of deep learning has facilitated research on plant point cloud segmentation. However, previous research has mainly focused on methods based on hard voxelization or down sampling. These methods are limited to segmenting simple plant organs. Segmentation of complex plant point clouds with high spatial resolution remains challenging. Recent deep learning methods to segment the to point clouds have emerged to address this challenge. They learn features from the input data in a data-driven manner [23]. Thanks to advanced neural networks, deep learning has shown great potential in plant 3D phenotypic analysis, surpassing most traditional segmentation methods (Guo et al., 2020). However, high-precision 3D point clouds places high demands on computer performance, especially graphics cards (Shi et al., 2019). Ghahremani et al. proposed a new pattern-based deep neural network, Pattern-Net, for segmentation of wheat point clouds. For the first time, they partitioned wheat point clouds into defined organs. And their features are analyzed directly in 3D space. The network provides a method for the actual segmentation of plant parts directly in the point cloud domain (Ghahremani et al., 2021). However, the network requires high computer performance and average recognition accuracy. A point-based fully convolutional neural network (PFCN) was proposed by Jin et al. The network directly uses points containing only geometric information. Then extracts point-by-point and block-by-block features to classify each point. The method addresses the difficulty

of segmenting large-scale forest scenes (Jin et al., 2020). A bifunctional deep learning neural network (PlantNet) was proposed by Li et al. The method implemented semantic segmentation and instance segmentation of two dicotyledons and one monocotyledon in a point cloud (Li et al., 2022). However, high-precision 3D point cloud processing places high demands on computer performance, especially on the size of graphics memory. Therefore, existing network architectures and hyperparameters are mainly designed for small-scale inputs under current hardware limitations. Down sampling the point clouds before feeding it into a point-based network is necessary, but it also leads to significant loss of original data information.

In summary, the main challenge for segmentation of dense and small sized rapeseed siliques in two reasons. Firstly, both the hard voxelization, which is widely used in voxel-based methods, and the down sampling operation in point-based methods result in a significant loss of information from the original data. Secondly, training and inferring on dense pixel grids or point clouds may impose intolerable computational costs on existing deep learning methods. Therefore, this paper proposes an efficient and high-precision deep learning segmentation method for sparse point clouds that uses DGCNN (Wang et al., 2019) to segment and then maps the results onto dense point clouds. This method accurately identifies and segments rapeseed siliques while maintaining complete spatial information and requiring low computer performance, making it a low-cost and high-efficiency method for silique identification. The method proposed in this paper provides a useful example of deep learning applied to dense 3D point clouds segmentation and lays the foundation for low-cost and high-efficiency intelligent measurement of rapeseed, promoting further research on rapeseed.

2 Materials and methods

2.1 An overview of the proposed silique identification methods

First, a circle video was captured around the rapeseed plant using a smartphone. Then the frames are filtered to meet the requirements of high definition, a complete view and no significant redundancy. Some of the over-exposed and dark images are then enhanced to obtain a high-quality video collection to improve the speed and quality of the 3D reconstruction. The processed video frames are then used as input to obtain a 3D dense point cloud after sparse reconstruction, depth estimation, and dense reconstruction (Figure 1A).

Next, the whole rapeseed silique point cloud was then obtained using straight pass filtering and down sampling to obtain the target horn fruit point cloud. The DGCNN network segmentation method was then used on this target point cloud. The down-sampled canopy canola rapeseed silique with the stalks removed were obtained (Figure 1B).

Then, the sparse-dense point cloud mapping segmentation method was then applied to the down sampled canopy rapeseed silique to obtain the original canopy rapeseed silique with stalks removed (Figure 1C).

Finally, Euclidean clustering was used to segment for canopy siliques. For the adherent siliques a linear fitting method based on a random sampling consistency algorithm was used to segment the identified siliques (Figure 1D).

The overall flowchart is shown in Figure 1.

2.2 Experimental materials and data collection

Mature rapeseed plants, whose varieties are Zhongshuang 6, Dadi 55, and Huayouza 62, were collected from experimental fields in Ezhou Base in May 2022. Thirty plants each variety were manually counted for the number of siliques per plant. Image acquisition was conducted indoors under natural light. The entire rapeseed plant was fixed in a flowerpot and photographed using a smartphone. The rapeseed was placed in the center of a device with a black background cloth surrounding it. A Xiaomi 10 smartphone was used to record 4K (3840 x 2160, 30fps) videos, with default imaging mode selected. The distance between the phone and rapeseed siliques was about 1-2 meters, and the video was about 30 seconds long. The distance between the phone and the siliques of the rapeseed depends on the size of the rapeseed. The larger the rapeseed the greater the distance. Simply fill your phone's viewfinder with the entire rapeseed. Some of the captured video images are shown in Figure 1A. The algorithm development and testing platform for this paper was a general-purpose computer (Intel Core 11th generation i9-12900K processor with a frequency of 2.5-5.2 GHz, 64GB memory, NVIDIA GeForce RTX 3090 graphics card with 24GB graphics memory), Windows 10 Professional Edition, VS 2019 + PCL 1.80, and Python 3.6.

2.3 3D reconstruction based on video key frames and SFM

Using sequential images, the steps for 3D reconstruction include image data collection and preprocessing, sparse point clouds reconstruction, depth map estimation, and dense point clouds reconstruction. The reconstruction process is shown in Figure 1A. SFM method requires capturing 100 images from different angles, which is tedious and inefficient. Using a smartphone to capture short videos is undoubtedly a more efficient and effortless way. However, compared to using a single image sequence, 3D reconstruction, based on video and SFM, requires solving two new problems. Firstly, 3D reconstruction, based on SFM, requires feature point extraction and matching. However, the unavoidable shaking during smartphone shooting can cause image blurring and dragging, making feature point detection

10.3389/fpls.2023.1188286

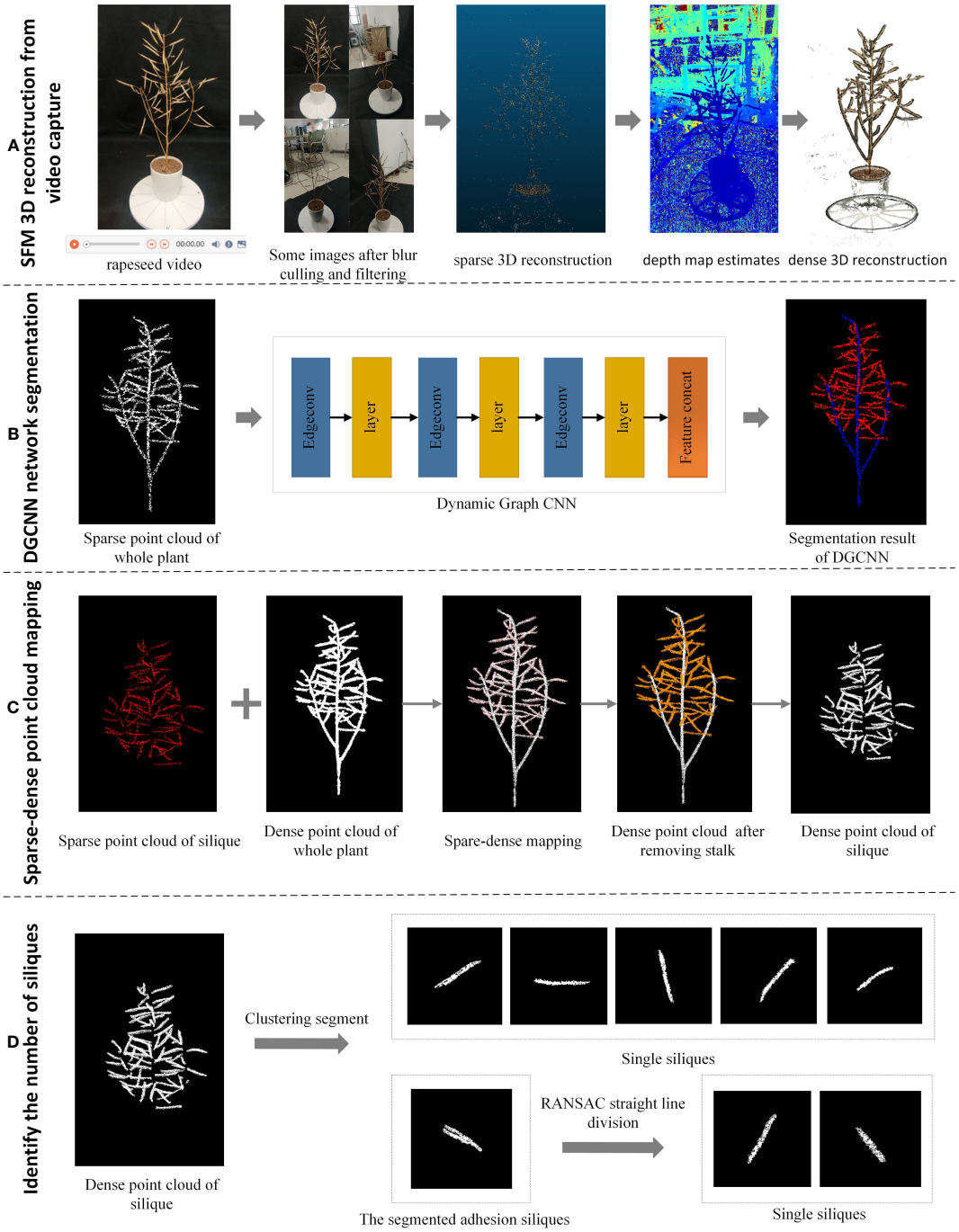


FIGURE 1
Flow chart of overall silique identification and counting. **(A)** Process for SFM 3D reconstruction from video capture **(B)** Segmentation of canopy silique point clouds after downsampling using the DGCNN network **(C)** A segmentation method using sparse-dense point cloud mapping to obtain the original canopy silique process **(D)** A straight-line fitting approach to segmentation using the Euclidean clustering-RANSAC algorithm to identify siliques process.

difficult. Secondly, SFM algorithm only requires around 100 sequence images. But a 30-second video captured by a smartphone can have 900 frames after decompression. Hence, directly using these images for 3D reconstruction will consume enormous computing power (Pepe et al., 2022). To address these two problems, we propose targeted solutions. On the one hand, for blurred images, Laplacian blurred image rejection based on convolutional variance can eliminate them. On the other hand, for redundant images, we use a similar image filtering method based on feature-matching similarity.

2.3.1 Laplacian blurred image rejection based on convolutional variance

A Laplacian convolution is performed on the image to calculate the variance, which is used to measure the image clarity. The calculation result quantifies the high-frequency information in the image. The larger the numerical value, the more high-frequency information the image contains, corresponding to a clear image; while a smaller numerical value corresponds to a blurry image with lost edge details. The commonly used Laplacian template has the disadvantage of weak anti-noise interference ability, so it needs to be smoothed before convolution. Bilateral filtering can effectively remove noise while preserving edges. In this paper, we choose to first apply bilateral filtering to the image for denoising, then convert the color image to grayscale, and perform Laplacian convolution to obtain the final variance (Rehman et al., 2022). This can be expressed mathematically as:

$$F = \frac{\sum_{i=0}^{I}\sum_{j=0}^{J}(P(\mathrm{i,j}) - \overline{P})^2}{I*J} \quad (1)$$

Where F is the sharpness value, I and J are the length and width of the 2D image, $P(i, j)$ is the pixel value at a point of the image and P is the average pixel value of the 2D image.

2.3.2 Similar image filtering based on feature-matching similarity

To extract a subset of keyframes from a large number of video frames while avoiding including too many redundant images, similarity detection on the images is necessary. The image similarity calculation based on feature point matching can solve this problem. For the same object, even if there is a significant change in angle between two images, our method can still judge the similarity based on the number of matched feature point pairs detected. The results obtained are relatively accurate. Each normal image has a certain number of feature points distributed throughout the image, such as edge points and corner points. Matching the feature points detected from two images, the higher the number of correctly matched pairs, the higher the image similarity.

The first step: Read in the two images to be detected, img1 and img2, and convert them to grayscale images, gray1 and gray2.

The second step: Create a SURF feature extractor and select the fast library for approximate nearest neighbors.

The third step: Extract the image features of gray1 and gray2, namely the feature points and the feature vectors around the feature points.

The fourth step: Perform SURF feature matching on the extracted feature points, obtain the number of matches, Matches, and set an empirical threshold of 0.7 to remove points that do not meet the matching requirements, leaving the number of correctly matched features, MatchNum.

The fifth step: Calculate the ratio of the number of correctly matched feature points to the total number of feature points, which represents the similarity value, denoted by Similarity. The larger the Similarity value, the more similar the images are. The formula for calculating Similarity is:

$$Similarity = \frac{MatchNum * 100}{Matches} \quad (2)$$

The process of the video keyframe extraction algorithm is as follows: First, set the sampling interval K according to the video duration. Then take out the first frame image and check if it is a blurry image. If it is blurry, continue to check the next frame until it meets the requirements and is used as the starting frame. Then take out the frame at a distance interval of K and continue to check if it is a blurry image. After it meets the clarity requirements, calculate the similarity between these two frames. If the similarity reaches S1 but does not exceed S2, take out this frame image. If the similarity is higher than S2, detect the (K+I)th frame. If the similarity is lower than S1, detect the (K-I)-th frame until it meets the requirements. Repeat the above steps until all images have been checked.

2.4 Point clouds segmentation methods of rapeseed siliques

2.4.1 Deep learning dataset

First, the 3D model of rapeseed siliques was reconstructed based on SFM algorithm. Then, the Cloud Compare open-source toolbox was used to create rapeseed silique point clouds dataset by labeling the stems and canopy siliques of the training data separately. The stem part was labeled as 0, and the canopy siliques were named as 1, and saved as "txt" format. When training the model using deep learning methods, a sufficient number of training sample images are required. Consider the rotational translation invariance and scale invariance of a point cloud. For each point in the point cloud dataset, a random translation vector can be generated in the x, y and z directions. The point coordinates are then added to this vector to achieve a random translation. Each component of the translation vector can be randomly generated between [-0.2, 0.2]. Three scaling factors can also be generated randomly between [0.65, 1.7], corresponding to the scaling factors in the x, y and z directions respectively. The coordinates of the point are then multiplied by each of the three scaling factors. This results in random anisotropic scaling to augment the training data. A total of 90 sets of training datasets were created, which expanded the dataset to six times its original size. In total, 540 sets of rapeseed siliques point clouds training data were obtained. The training set was trained, validated, and tested with semantic segmentation models at 70%, 15%, and 15%.

2.4.2 Silique and stalk point clouds segmentation based on DGCNN

Dynamic Graph Convolutional Neural Network (DGCNN) is a model based on Graph Convolutional Neural Network (GCN). It is specifically used for tasks such as image classification, point cloud classification and semantic segmentation. DGCNN uses local neighborhood information to perform feature extraction on each node. The global information is captured by a graph convolutional neural network to enable classification and segmentation tasks. It uses an integrated convolution module, EdgeConv, as its core,

which models the points in the point clouds by using a graph approach. This enables the network to learn both local and global features of the point clouds while also learning the independent information of each point (Liu et al., 2022). Traditional 2D image convolution defines the local region of pixels using the size of the convolution kernel. EdgeConv, on the other hand, constructs a local region using a k-nearest neighbor (kNN) graph and performs convolution operations on it.

Since 3D point clouds are unstructured and unordered, most deep learning methods for processing 3D data in point clouds segmentation tasks convert the point clouds into a collection of sequential images or a voxel-based 3D data representation. However, multi-view and voxel-based representations can lead to unnecessary data redundancy and limit output resolution. Subsequently, PointNet (Qi et al., 2017a) directly processes 3D point clouds as deep neural network input data, but only based on global features of the point clouds, lacking local features. Therefore, PointNet++ (Qi et al., 2017b) proposes grouping and layering the point clouds, using PointNet to capture both local and global information. However, this method lacks the association between points. The introduction of DGCNN, based on PointNet++, adds the relationship between points, making local information more prominent. Compared to traditional CNN models, DGCNN is able to handle unstructured data and capture the relationship between local and global features. DGCNN can also be used to classify and segment unordered point sets without considering the order in which the points are arranged. In addition, DGCNN also has better robustness and is more capable of handling noisy and incomplete data. For the segmentation of the entire rapeseed siliques, high-precision local information is required for the segmentation of the siliques and stems, which can be met by DGCNN. Therefore, in this paper, DGCNN is used for semantic segmentation of rapeseed images to obtain the point clouds of the siliques without the stems.

This network model architecture is used as a model architecture for classification (top branch) and segmentation (bottom branch). The classification model takes as input n points, computes an edge feature set of size k for each point at the EdgeConv layer, and aggregates within each feature set to compute the response of the EdgeConv counterpart of the corresponding point. The output features of the last EdgeConv layer are aggregated globally to form a one-dimensional global descriptor that is used to generate the classification scores for class c. The partitioning model extends the classification model by connecting the one-dimensional global descriptor to all EdgeConv outputs. The EdgeConv output (as a local descriptor). It outputs a classification score for each point for the p semantic labels. The point cloud transformation block aims to align the input point set to the typical space by applying an estimated 3 × 3 matrix. To estimate the 3 × 3 matrix, a tensor is used that connects the coordinates of each point with the difference in coordinates between its k neighbors. The EdgeConv module takes as input a tensor of shape $n \times f$ and computes the edge features of each point by applying a multilayer perceptron (mlp) with the number of layer neurons defined as $\{a_1, a_2, \ldots, a_n\}$, and generating a shape tensor $\times a_n$ after the set of neighboring edge features.

Since the network provides the probability prediction for each point in each class, the maximum probability value of the class and the point clouds label are used together to calculate the loss during network training. The network parameters are trained and learned through backpropagation. During network training, point clouds are input in batches (batch size) to reduce their differences in type and geometry. The equation of loss calculation during the training process uses the classic cross-entropy loss function as shown in (3).

$$LOSS(x, label) = -\log\left(\exp[x(label)] / \sum j \exp[x(j)]\right) \quad (3)$$

where x is the output of the network, the label is the corresponding label and j is the order of the output.

During the training process, the network needs to learn how to set parameters based on the results of the loss function calculation in each epoch, which is directly influenced by the manually labeled data. The hyperparameters that need to be manually set in the network (parameters that cannot be learned) control the speed of convergence of the network's loss and the training effect. This is because the more the loss calculation results converge, the better the semantic segmentation effect of the network. The hyperparameter settings in this article mainly focus on the optimizer, learning rate, training epochs, and batch size of inputs in each epoch. The optimizer used is Adam; the learning rate is set to 0.001; the training epoch value is set to 200; and the batch size is set to 20. The new point cloud data is segmented using a trained DGCNN network. The segmentation process consists of two processes: forward propagation and backward propagation. The forward propagation process is to input the point cloud data into the DGCNN network and get the output result of the network. The backward propagation process is to calculate the gradient of the loss function based on the network output result. The back propagation algorithm is used to update the network parameters, thus enabling the network to segment the point cloud data better. These parameters represent the maximum computational capacity supported by the computer during experimentation. The segmentation results are shown in Figure 2.

2.4.3 Kd-tree radius searches - a segmentation method for sparse-dense point clouds mapping

Point clouds processing requires extremely high computational power. The DGCNN can effectively segment rapeseed point clouds. However, the computer used in this study is unable to process point clouds with more than 16,384 points. To handle large rapeseed point clouds, a new method of sparse-dense point clouds mapping is needed to solve the bottleneck caused by insufficient computational power. This method, based on the sparse crown siliques obtained by DGCNN segmentation, uses Kd-tree radius search to obtain the original dense point clouds of crown siliques (Evangelou et al., 2021).

KdTree is a high-dimensional space indexing structure used to partition k-dimensional data space. Its essence is a binary search tree with constraints. Approximate search algorithms based on Kd-Tree can quickly and accurately find the nearest neighbors of a search point, which is often used in feature point matching based on similarity (Xiao et al., 2022). There are two basic methods for similarity search algorithms in index structures: one is radius searches, and the other is K-neighbor searches (Zhang et al.,

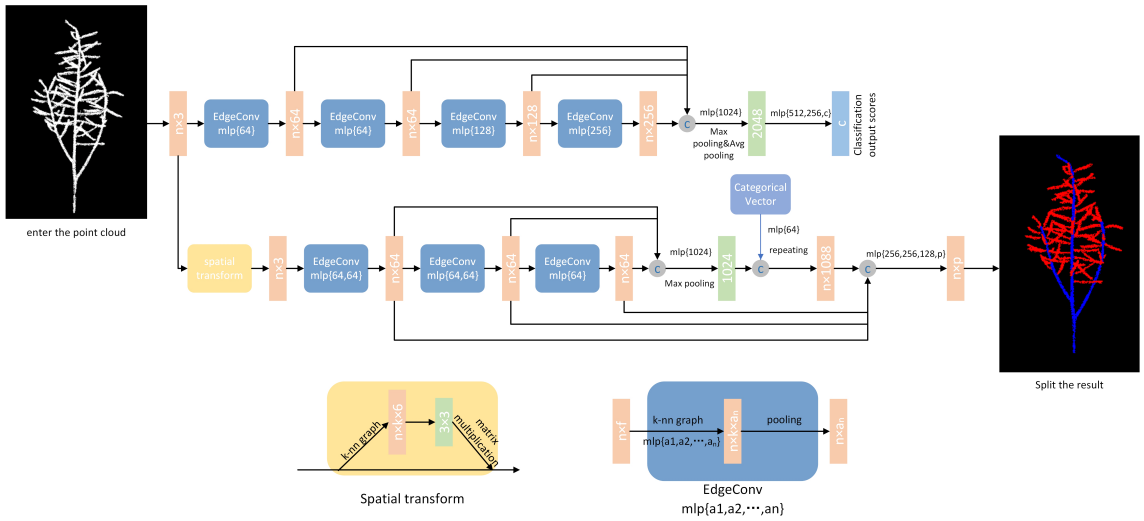


FIGURE 2
DGCNN network structure and segmentation effect. Model architectures: The model architectures used for classification (top branch) and segmentation (bottom branch). The classification model takes as input n points, calculates an edge feature set of size k for each point at an EdgeConv layer, and aggregates features within each set to compute EdgeConv responses for corresponding points. The output features of the last EdgeConv layer are aggregated globally to form an 1D global descriptor, which is used to generate classification scores for c classes. The segmentation model extends the classification model by concatenating the 1D global descriptor and all the EdgeConv outputs (serving as local descriptors) for each point. It outputs per-point classification scores for p semantic labels. ©: concatenation. Point cloud transform block: The point cloud transform block is designed to align an input point set to a canonical space by applying an estimated 3×3 matrix. To estimate the 3×3 matrix, a tensor concatenating the coordinates of each point and the coordinate differences between its k neighboring points is used. EdgeConv block: The EdgeConv block takes as input a tensor of shape n×f, computes edge features for each point by applying a multi-layer perceptron(mlp) with the number of layer neurons defined as $\{a_1, a_2,..., a_n\}$, and generates a tensor of shape $n \times a_n$ after pooling among neighboring edge features.

2020). Radius searches means finding all data in the dataset that is within the given search distance threshold (with the search point as the center and the search distance as the radius) that is less than the threshold distance from the search point (data within the radius). K-nearest neighbor searches is finding the K closest data points to the search point from the dataset. When K=1, it becomes the nearest neighbor searches. For 3D point clouds, all K-D trees are 3D. Building a k-d tree is a recursively unfolding process: at each level of expansion, all remaining datasets are divided along a specific dimension using a hyperplane perpendicular to the corresponding axis. At the root of the Kd-tree, all data is split according to the first dimension. The next level in the Kd-tree is divided along the next dimension. When all other dimensions are exhausted, it returns to the first dimension. The most efficient way to construct a K-D tree is to use a partitioning method similar to quicksort, placing the median at the root node, then placing values smaller than the median in the left subtree and values larger than the median in the right subtree, and finally repeating this process on the left and right subtrees until the last element is partitioned.

Figure 3 shows the process of removing the stems from rapeseed during the silique stage. The sparse crown siliques segmented by the DGCNN are shown in (a), while (b) shows the original point clouds

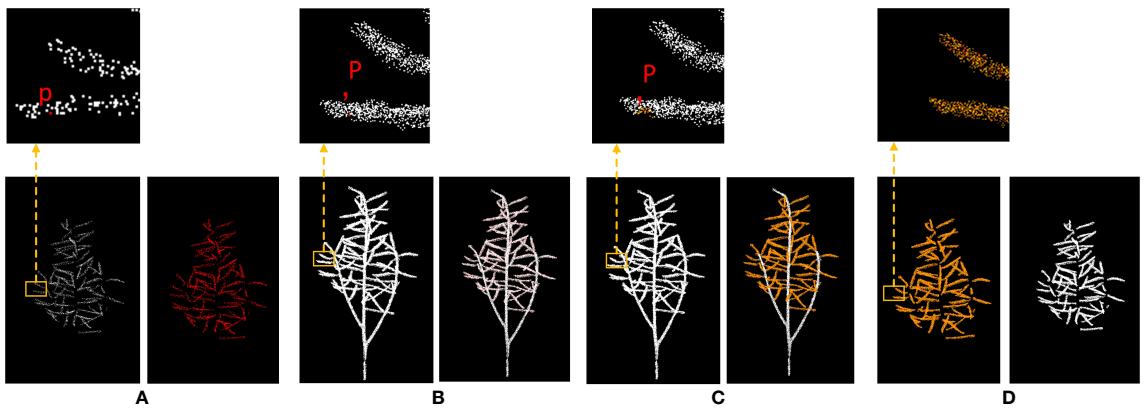


FIGURE 3
Diagram of the process of removing stalks from rapeseed silique. **(A)** Canopy silique point cloud after sampling is desired **(B)** Diagram of the original carob point cloud index process **(C)** Diagram of the search process for raw carob point clouds **(D)** Raw rapeseed canopy mirage point cloud.

of the entire rapeseed plant. The upper portion of (a) is a magnified view of the sparse crown siliques segmented by the DGCNN, while the upper portion of (b) shows the original point clouds of the entire rapeseed plant. In (a), a point is selected and its corresponding points in (b) is found. point clouds with a radius of 0.01 is then retained around this point. The retained point clouds are shown in yellow in (c). The final segmentation result is a dense crown silique, as shown in (d), which prepares for the accurate segmentation of individual silique in the subsequent steps.

2.5 Rapeseed silique identification based on the RANSAC algorithm

The contour of the reconstructed 3D point clouds of the rapeseed siliques is complete. Therefore, it is more suitable for line fitting identification. In this study, the plant point clouds is first segmented using the Euclidean clustering (Chen and Zhang, 2004). Then, based on the RANSAC algorithm (Barath et al., 2022), straight line fitting is applied to the clustered point clouds to segment the point clouds of the rapeseed siliques. Finally, the identification results are counted.

Plant point clouds exist in the form of multiple point clouds clusters in 3D space. In order to effectively detect and count the number of rapeseed siliques, it is necessary to determine the number of point clouds clusters and process each cluster separately to avoid missing or mis-segmenting the identification. Therefore, the plant point clouds are first segmented into multiple point clouds clusters using the distance-based Euclidean clustering algorithm, and then each cluster is processed separately. The point clouds clusters can be divided into two categories: those containing only one rapeseed silique point clouds, and those containing two or more rapeseed silique point clouds. Straight line segmentation is performed on each cluster. If the cluster contains rapeseed silique point clouds, the corresponding straight line can be fitted. Finally, the number of straight lines fitted for all point clouds clusters is counted to obtain the number of rapeseed siliques in the entire plant. A point clouds cluster may contain multiple rapeseed siliques, so the straight-line point clouds is fitted and segmented multiple times until the remaining un-fitted point clouds in the cluster is less than a certain value. Based on the measured width of the rapeseed siliques, the width range of the straight-line model is set to 0.03-0.05m, and the error threshold between the inliers and the model is set to 0.018m. At least 200 points are required to segment a straight-line model. During the straight-line fitting process, if the number of points in the point clouds is less than 200, the current model fitting result is regarded as a misidentification result. The straight-line model obtained in the previous fitting is retained as the identification result, and the current fitting result is discarded.

3 Experimental results and analysis

3.1 3D reconstruction experiments based on video and SFM

An experiment was designed to compare three methods of 3D reconstruction: keyframe, sequence image, and fixed. Keyframes refer to the frames that capture the critical movements or changes of an object. Sequence images are the pictures taken around the rapeseed siliques, while fixed-frames are the video frames captured around the rapeseed siliques, with one frame extracted every ten frames. The experiment ensured that both sequence images and fixed-frames had 90 input images, while the keyframes were determined based on the algorithm results from the previous steps. The shooting method for sequence images was consistent with that of video frames. The quality of the reconstruction was compared from two aspects: the number of

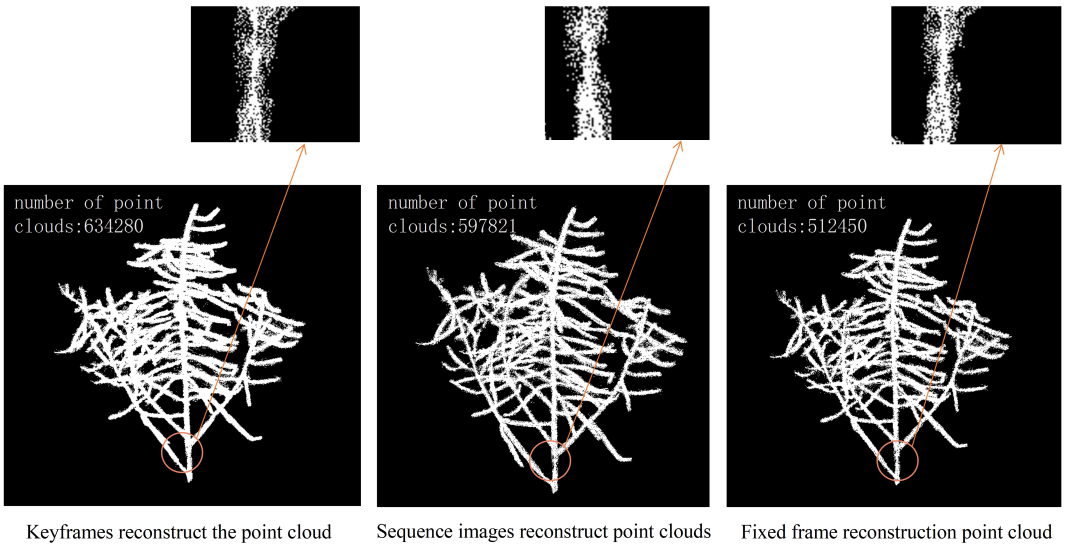


FIGURE 4
Comparison of point cloud quality obtained by three reconstruction methods.

point clouds and the degree of restoration of details. That is, for the same rapeseed plant, the more point clouds, the more complete the structure and the fewer holes and missing parts. At the same time, the details of the point clouds, such as the siliques and stems, were observed, and the reconstruction effect of the details represented the quality of the point clouds. Figure 4 shows three point clouds obtained 3D reconstruction using keyframe, sequence image and fixed frame, respectively, from left to right. The small images below Figure 4 are an enlarged display of the circled part, which facilitates a more intuitive analysis of the reconstruction quality. Upon observing the enlarged part, it was found that the three methods had different reconstruction degrees of the 3D stem structure, with the keyframe reconstructing the most complete stem details, followed by the sequence image, and the fixed frame having the worst reconstruction. Through the analysis and comparison of point clouds obtained from multiple 3D reconstruction, it was found that the algorithm proposed in this paper had significant improvements compared to 3D reconstruction based on fixed frame, which could greatly improve the accuracy of subsequent phenotypic measurements of rapeseed siliques.

We conducted experiments on 12 rapeseed plants from three different varieties. The required time for 3D reconstruction based on the three methods was recorded and compared. The comparison of time consumption for these three methods is shown in Figure 5. For the same rapeseed plant, the duration of sequence image-based 3D reconstruction was significantly higher than that of fixed-frame and keyframe. 3D reconstruction based on fixed frame and keyframe, had similar durations, but keyframe-based 3D reconstruction was slightly shorter than fixed-frame-based reconstruction. Fixed-frame-based reconstruction improved the efficiency of reconstruction by 19.39% compared to sequence image-based reconstruction, while keyframe-based reconstruction improved efficiency by 24.4% compared to sequence image-based reconstruction. Therefore, if judged solely on the basis of the reconstruction time, sequence image-based reconstruction took the longest time, followed by fixed-frame-based reconstruction, and keyframe-based reconstruction took the shortest time.

3.2 Experimental segmentation of silique point clouds

Semantic segmentation based on deep learning has a very broad development prospect in the field of computer vision. However, many network models with good segmentation results occupy a large amount of memory and take a long time to process 3D point clouds (Mirande et al., 2022). Based on the DGCNN-sparse-dense point clouds mapping, it has faster processing speed and better segmentation results, and consumes less memory. We conducted experiments on four rapeseed siliques. Figure 6 shows the segmentation results of PointNet, PointNet++, and DGCNN for the four rapeseed siliques, respectively.

Table 1 shows a quantitative comparison of the three networks. DGCNN achieved the best results in most cases and outperformed other networks in all four average quantization metrics.

The reconstructed rapeseed silique point clouds were manually segmented into rapeseed stems and canopy siliques using CloudCompare software. The segmented point clouds were down

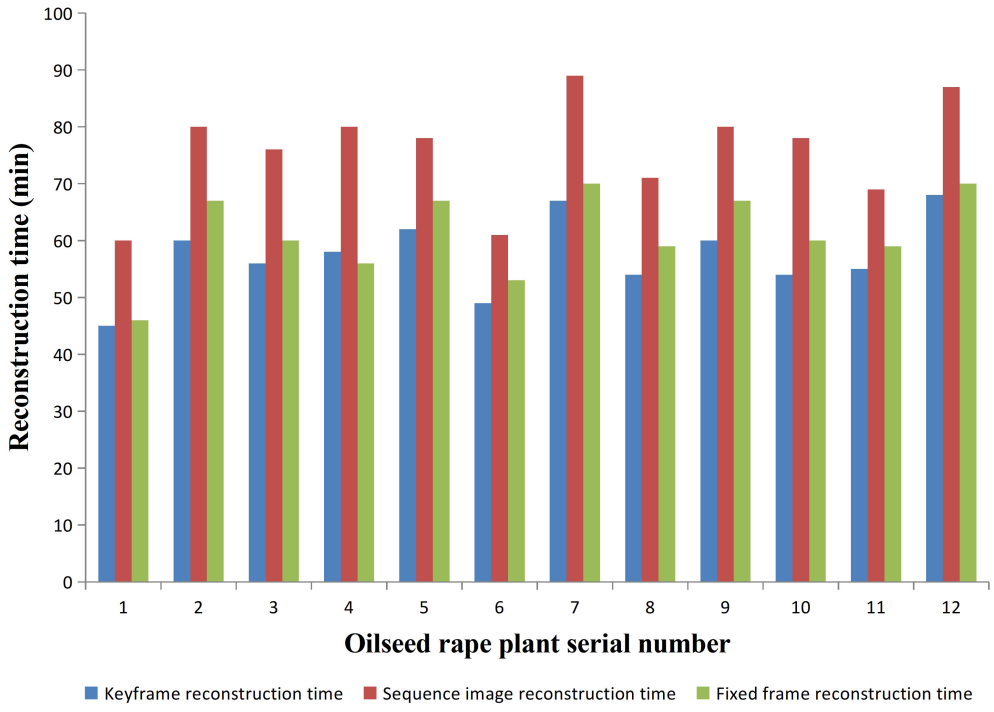


FIGURE 5
Comparison of reconstruction time of three reconstruction methods.

10.3389/fpls.2023.1188286

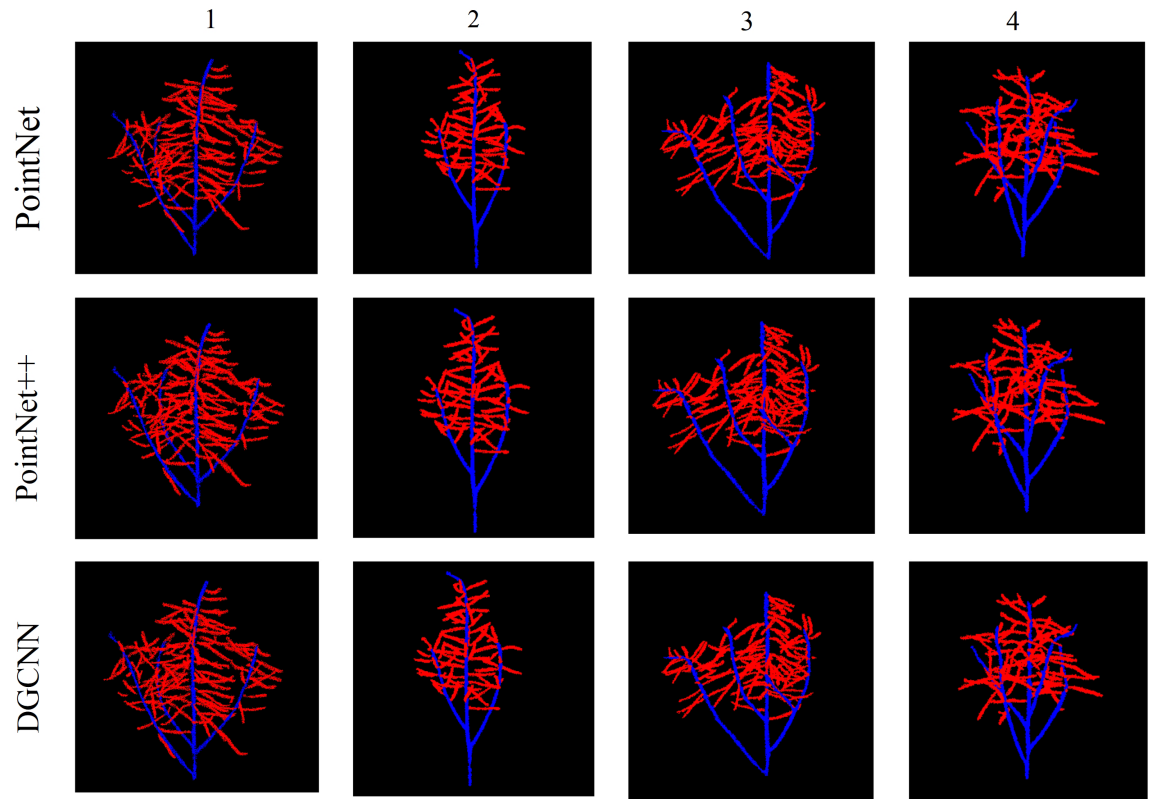


FIGURE 6
Renderings of three types of network segmentation.

TABLE 1 The semantic segmentation performance of the three networks was quantitatively compared.

	Methods	1	2	3	4
Prec (%)	PointNet	77.16	93.56	89.18	96.61
	PointNet++	87.79	95.91	91.71	96.70
	DGCNN	**92.72**	**96.37**	**92.39**	**97.97**
Rec (%)	PointNet	91.21	89.08	86.89	98.28
	PointNet++	90.84	83.96	91.94	97.82
	DGCNN	**97.7**	**89.55**	**93.32**	**98.21**
F1 (%)	PointNet	93.67	86.87	87.87	97.76
	PointNet++	94.84	91.92	91.82	97.79
	DGCNN	**97.25**	**93.86**	**94.54**	**98.31**
IoU (%)	PointNet	88.09	92.42	93.42	93.26
	PointNet++	90.18	96.45	96.65	95.51
	DGCNN	**94.35**	**95.68**	**96.77**	**96.74**

TABLE 2 Quantitative comparison of silique down sampling to different point cloud numbers.

The number of point clouds after down sampling	4096	8192	12288	16384	18432
Times(h)	1.25	2.5	3.75	5	×
The number of point clouds after restoration	×	250241	272783	301013	×
Silique recognition rate	×	98.12%	91.99%	98.03%	×

sampled to 4096, 8192, 12288, 16384, and 18432 points for each rapeseed siliques. Due to the large number and complex structure of the rapeseed silique point clouds, down sampling to 4096 points not only reduced the number of point clouds too much but also destroyed the structure. When the number of point clouds is too small, it cannot be restored to the original number of point clouds under the method of sparse-dense point clouds mapping, which leads to the inability to recognize subsequent siliques. On the other hand, down sampling to 18432 requires a lot of resources, making it impossible for the computer to run. Table 2 shows that when down sampled to 8192 point clouds, the training and processing time is the shortest and the siliques recognition rate is the highest after sparse-dense point clouds mapping.

3.3 Experimental recognition of silique point clouds

The recognition results were compared with the ground truth. As shown in Table 3, numbers 1-4 were Huayouza 62, numbers 5-8 were Zhongshuang 6, and numbers 9-12 were Dadi 55. The total recognition precision was 97.80%, the mean absolute percentage error was 1.96% and the R^2 was 0.96. Figure 7 shows the segmentation results of the siliques. Figures 7A, E show two different cases of siliques adhesion. Figures 7C, B ,D, F, G show the recognition results under the adjacent or adhesive state of the siliques. Even when the siliques were heavily adjoined and occluded, our method could distinguish and recognize different siliques. Our method segmented and recognized the siliques based on their spatial shape features. However, there were a small number of missed recognitions during the silique's recognition process. The reasons for missed recognition were mainly due to the silique point clouds having too few points or irregular shapes, which did not meet the fitting conditions set by the RANSAC algorithm. When the siliques were extremely small in shape and had fewer point clouds after down sampling and statistical filtering, they could be easily treated as noise and removed. In addition, incomplete or irregular straight-line contours resulted in the inability to fit a straight line.

4 Discussion

(1) In this paper the SFM method was chosen for the 3D reconstruction of rapeseed silique. The method is inexpensive and the quality of the 3D reconstruction is high, but the time consumption is huge. The time required for one 3D reconstruction was more than 30 minutes. Moreover, more images are required to obtain higher quality 3D point clouds, which leads to a dramatic increase in time consumption. Therefore, further research is needed to explore optimization methods for SFM, such as adding GPU acceleration.

(2) The point cloud segmentation method proposed in this paper requires down sampling of whole rapeseed plant point cloud. As can be seen from the above experiments, due to the large number of point clouds and complex structure of the whole rapeseed silique. When down sampling to 4096, not only the number of point clouds is too small, but also the structure is destroyed. When the number of point clouds is too small, it cannot be restored to the original number of point clouds under the sparse-dense point cloud mapping method. This in turn results in the subsequent number of siliques not being identified. And when down sampling 18432, a large number of resources are required causing the computer to become inoperable.

TABLE 3 Silique identification and counting results.

Rapeseed varieties	Plant number	True value	Correct identification number of siliques	Missing identification number of siliques	Pr (%)
Huayouza 62	1	126	124	2	98.41
	2	196	192	4	97.96
	3	110	108	2	98018
	4	119	117	2	98.32
Zhongshuang 6	5	104	102	2	98.08
	6	139	135	4	97.12
	7	85	83	2	97.65
	8	99	98	1	98.99
Dadi 55	9	129	126	3	97.67
	10	105	104	1	99.05
	11	124	121	3	97.58
	12	118	115	3	97.46
Total		1457	1425	32	97.80

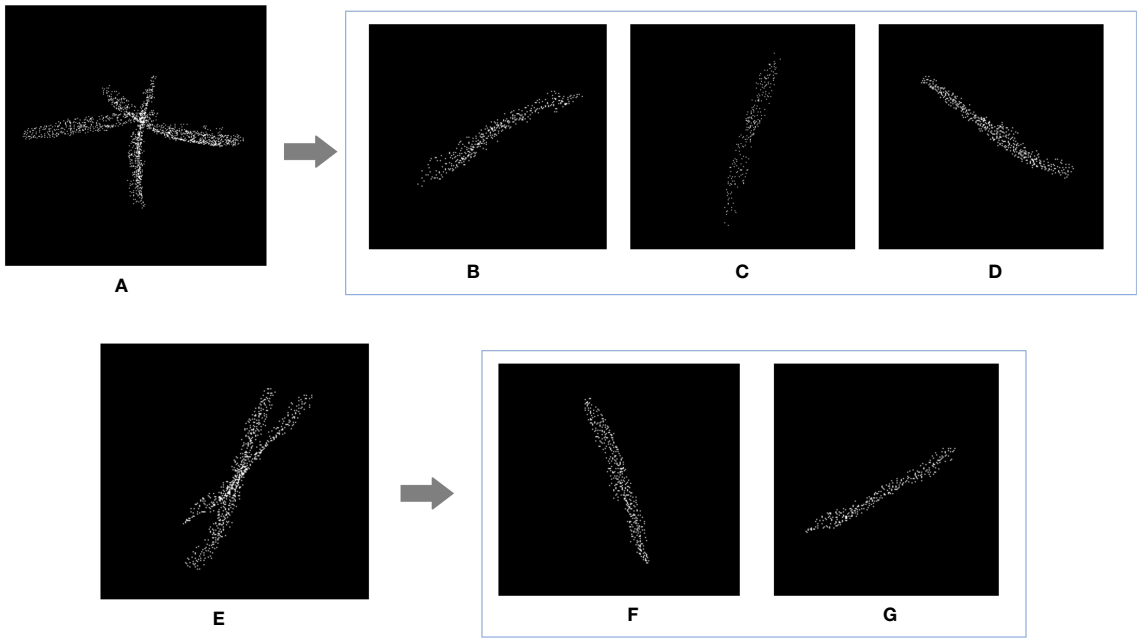


FIGURE 7
Rapeseed silique segmentation results. **(A)** Diagram of three siliques adhesions **(E)** Diagram of two siliques adhesions **(B, C, D, F, G)** Single silique.

Therefore, the range for down sampling the whole rapeseed silique point cloud is 4096-18432.

(3) The Euclidean clustering-RANSAC segmentation and identification method has certain limitations in the segmentation of the siliques point cloud identification. The number of siliques identified by this method is often slightly smaller than the actual number. There are three main reasons for this. Firstly, in the image pre-processing process, some small siliques are mistakenly rejected. Secondly, in the point cloud filtering session, some siliques were broken into small pieces and mistakenly considered as outlier noise and were rejected. Thirdly, the parameter settings for segmenting the point cloud of rapeseed silique based on the RANSAC algorithm for linear fitting of the clustered point cloud. If the width range of the line model is set to less than 0.03m or greater than 0.05m, it can cause changes in the siliques shape and inaccurate counting.

5 Conclusion

This paper provides a process and methodology that can be used as a reference for segmenting dense plant point clouds with complex structures. We take the dense point clouds of whole rapeseed plants at the siliqua stage with complex morphological features as a typical example. The DGCNN used in this paper performs semantic segmentation on the entire rapeseed point clouds. Compared with instance segmentation, the data annotation cost of semantic segmentation is much lower. Moreover, the DGCNN adds relationships between points on the basis of PointNet++, making local information more prominent. For the entire rapeseed at the siliqua stage, high-precision local information is required for the segmentation between siliqua and stem. DGCNN is well suited for this task. After segmentation, the sparse canopy siliqua point clouds without stem are obtained, and then the sparse-dense point clouds mapping segmentation method is used to segment the original rapeseed canopy siliqua point clouds. This method greatly reduces the computational requirements for deep learning network point clouds segmentation. Targeted solutions are proposed for several difficulties in identifying mature rapeseed siliqua.

(1) This paper proposes a similarity-based video keyframe extraction algorithm. The algorithm effectively removes redundant and motion-blurred images, saving reconstruction time and cost, and ultimately ensuring a more complete view of the extracted video frames. Image enhancement processing of video frames improves contrast and enhances edge details, thereby improving the quality of the 3D model.

(2) To solve the problem of recognizing and counting the overall number of rapeseed siliques, this paper proposes a method based on DGCNN-sparse-dense point clouds mapping to segment the crown siliques. This method can remove the stems of the entire rapeseed siliques while maintaining complete spatial information, compared to only using DGCNN network to segment sparse crown siliques. This method not only improves the accuracy of subsequent siliques point clouds recognition and counting, but also greatly reduces the algorithm's requirements for computer computing power.

(3) For the crown siliques, this paper uses Euclidean clustering segmentation and a line fitting method based on random sample consensus algorithm to segment and recognize the siliques. Based on the contour features of line shapes, this method can greatly improve the accuracy of siliques recognition, which is of great significance for yield estimation and subsequent cultivation.

(4) This paper identifies 1457 siliques from 12 rapeseed plants, with a total identification accuracy rate of 97.80%. When comparing the total calculated number of siliques with the actual value, the coefficient of determination is 0.97, and the average absolute percentage error is 1.96%. This method can effectively recognize adhered siliques, as well as identify and count the entire siliques. The proposed method not only has extremely low cost, good portability, and high precision, but also can effectively save runtime and improve efficiency, greatly improving the accuracy of siliques counting.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

YQ: Conceptualization, lnvestigation, Methodology, Visualization, Writing-original draft. MZ: Supervision, Software. ZH: Data curation. SW: Data curation. BH: Formal analysis. CP: Validation. QL: Funding acquisition, Supervision, Resources. SX: Methodology, Investigation, Supervision,Writing-review editing. GZ: Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the National Key Research and Development Program of China (Program No. 2018YFD1000904) and the Fundamental Research Funds for the Central Universities (2662022GXYJ001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Barath, D., Cavalli, L., and Pollefeys, M. (2022). "Learning to find good models in RANSAC," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (New Orleans, LA, USA: IEEE/CVF) 15744–15753. doi: 10.1109/TSMCB.2004.831165

Chen, S., and Zhang, D. (2004). Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Trans. Systems Man Cybernetics Part B (Cybernetics)* 34 (4), 1907–1916. doi: 10.1109/TSMCB.2004.831165

Dandrifosse, S., Ennadifi, E., Carlier, A., Gosselin, B., Dumont, B., and Mercatoris, B. (2022). Deep learning for wheat ear segmentation and ear density measurement: from heading to maturity. *Comput. Electron. Agric.* 199, 107161. doi: 10.1016/j.compag.2022.107161

Du, R., Ma, Z., Xie, P., He, Y., and Cen, H. (2023). PST: Plant segmentation transformer for 3D point clouds of rapeseed plants at the podding stage. *ISPRS J. Photogrammetry Remote Sens.* 195, 380–392. doi: 10.1016/j.isprsjprs.2022.11.022

Evangelou, I., Papaioannou, G., Vardis, K., and Vasilakis, A. (2021). Fast radius search exploiting ray-tracing frameworks. *J. Comput. Graphics Techniques* 10 (1).

Ghahremani, M., Williams, K., Corke, F. M., Tiddeman, B., Liu, Y., and Doonan, J. H. (2021). Deep segmentation of point clouds of wheat. *Front. Plant Sci.* 12, 608732. doi: 10.3389/fpls.2021.608732

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M. (2020). Deep learning for 3d point clouds: a survey. *IEEE Trans. Pattern Anal. Mach. Intell (TPAMI)*, 42(4), 994–1011. doi: 10.1109/TPAMI.2020.3005434

Jin, S., Su, Y., Zhao, X., Hu, T., and Guo, Q. (2020). A point-based fully convolutional neural network for airborne LiDAR ground point filtering in forested environments. *IEEE J. selected topics Appl. Earth observations Remote Sens.* 13, 3958–3974. doi: 10.1109/JSTARS.2020.3008477

Jin, S., Sun, X., Wu, F., Su, Y., Li, Y., Song, S., et al. (2021). Lidar sheds new light on plant phenomics for plant breeding and management: recent advances and future prospects. *ISPRS J. Photogrammetry Remote Sens.* 171, 202–223. doi: 10.1016/j.isprsjprs.2020.11.006

Li, D., Shi, G., Li, J., Chen, Y., Zhang, S., Xiang, S., et al. (2022). PlantNet: a dual-function point cloud segmentation network for multiple plant species. *ISPRS J. Photogrammetry Remote Sens.* 184, 243–263. doi: 10.1016/j.isprsjprs.2022.01.007

Li, S., Yan, Z., Guo, Y., Su, X., Cao, Y., Jiang, B., et al. (2022). SPM-IS: an auto-algorithm to acquire a mature soybean phenotype based on instance segmentation. *Crop J.* 10 (5), 1412–1423. doi: 10.1016/j.cj.2021.05.014

Liu, Q., Dong, Y., Zhang, Y., and Luo, H. (2022). A fast dynamic graph convolutional network and CNN parallel network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 60(2), 1507–1520. doi: 10.1109/TGRS.2022.3179419

Liu, Z., Huang, W., and Wang, L. (2019). Automatic counting of wheat ears in field based on improved K-means clustering algorithm. *Trans. Chin. Soc Agric. Eng.* 35, 174–181. doi: 10.11975/j.issn.1002-6819.2019.04.021

Marzulli, M. I., Raumonen, P., Greco, R., Persia, M., and Tartarino, P. (2020). Estimating tree stem diameters and volume from smartphone photogrammetric point clouds. *Forestry: Int. J. For. Res.* 93 (3), 411–429. doi: 10.1093/forestry/cpz067

Mirande, K., Godin, C., Tisserand, M., Charlaix, J., Besnard, F., and Hétroy-Wheeler, F. (2022). A graph-based approach for simultaneous semantic and instance segmentation of plant 3D point clouds *IEEE Trans. Geosci. Remote Sens.* 60 (6), 4422–4433. doi: 10.3389/fpls.2022.1012669

Ni, X., Li, C., Jiang, H., and Takeda, F. (2021). Three-dimensional photogrammetry with deep learning instance segmentation to extract berry fruit harvestability traits. *ISPRS J. Photogrammetry Remote Sens.* 171, 297–309. doi: 10.1016/j.isprsjprs.2020.11.010

Pepe, M., Fregonese, L., and Crocetto, N. (2022). Use of SfM-MVS approach to nadir and oblique images generated throught aerial cameras to build 2.5 d map and 3D

models in urban areas. *Geocarto Int.* 37 (1), 120–141. doi: 10.1080/10106049.2019.1700558

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017a). "Pointnet: deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 652–660. doi: 10.1109/CVPR.2017.16

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017b). Pointnet++: deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.* 30. doi: 10.48550/arXiv.1706.02413

Rehman, S., Hussain, A., Rahman, J. U., Anjum, N., and Munir, T. (2022). Modified Laplace based variational iteration method for the mechanical vibrations and its applications. *Acta mechanica automatica* 16 (2), 98–102. doi: 10.2478/ama-2022-0012

Ren-feng, L., Shi-yao, H., Yong-peng, N., and Sheng-yong, X. (2020). Automated detection research for number and key phenotypic parameters of rapeseed silique. *Chin. J. Oil Crop Sci.* 42 (1), 71. doi: 10.19802/j.issn.1007-9084.2019092

Shengyong, X., Kun, L., Lili, P., Taige, L., Yuxin, Z., and Bo, W. (2019). "3D reconstruction of rape branch and pod recognition based on RGB-d camera," in *Nongye jixie Xuebao/Transactions of the Chinese society of agricultural machinery*, vol. 50(4), 273–280.

Shi, W., van de Zedde, R., Jiang, H., and Kootstra, G. (2019). Plant-part segmentation using deep learning and multi-view vision. *Biosyst. Eng.* 187, 81–95. doi: 10.1016/j.biosystemseng.2019.08.014

Shyam, P., Wu, X., and Bhat, S. (2012). History, evolution, and domestication of brassica crops. *Plant Breed. Rev.* 35, 19–84. doi: 10.1002/9781118100509

Tang, M., Tong, C., Liang, L., Du, C., Zhao, J., Xiao, L., et al. (2020). A recessive high-density pod mutant resource of brassica napus. *Plant Sci.* 293, 110411. doi: 10.1016/j.plantsci.2020.110411

Tao, Y., and Zhou, J. (2017). Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. *Comput. Electron. Agric.* 142, 388–396. doi: 10.1016/j.compag.2017.09.019

Wang, Y., Qin, Y., and Cui, J. (2021). Occlusion robust wheat ear counting algorithm based on deep learning. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.645899

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *ACM Trans. On Graphics (tog)* 38 (5), 1–12. doi: 10.1145/3326362

Xiao, P., Zhao, R., Li, D., Zeng, Z., Qi, S., and Yang, X. (2022). As-built inventory and deformation analysis of a high rockfill dam under construction with terrestrial laser scanning. *Sensors* 22 (2), 521. doi: 10.3390/s22020521

Zhang, W., Chen, X., Liu, Y., and Xi, Q. (2020). A distributed storage and computation k-nearest neighbor algorithm based cloud-edge computing for cyber-physical-social systems. *IEEE Access* 8, 50118–50130. doi: 10.1109/ACCESS.2020.2974764

Zhang, Y., Li, M., Ma, X., Wu, X., and Wang, Y. (2022). High-precision wheat head detection model based on one-stage network and GAN model. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.787852

Zhao, J., Kaga, A., Yamada, T., Komatsu, K., Hirata, K., Kikuchi, A., et al. (2023). Improved field-based soybean seed counting and localization with feature level considered. *Plant Phenomics* 5, 26. doi: 10.34133/plantphenomics.0026

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 July 2023
DOI 10.3389/fpls.2023.1163700

Check for updates

OPEN ACCESS

EDITED BY
Zhanyou Xu,
United States Department of Agriculture, United States

REVIEWED BY
Min Wang,
Nanjing Agricultural University, China
Karansher Singh Sandhu,
Bayer Crop Science, United States

*CORRESPONDENCE
Yitao Liao
liaoetao@mail.hzau.edu.cn

RECEIVED 11 February 2023
ACCEPTED 21 June 2023
PUBLISHED 18 July 2023

CITATION
Li M, Liao Y, Lu Z, Sun M and Lai H (2023) Non-destructive monitoring method for leaf area of *Brassica napus* based on image processing and deep learning. *Front. Plant Sci.* 14:1163700. doi: 10.3389/fpls.2023.1163700

COPYRIGHT
© 2023 Li, Liao, Lu, Sun and Lai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Non-destructive monitoring method for leaf area of *Brassica napus* based on image processing and deep learning

Mengcheng Li[1], Yitao Liao[1]*, Zhifeng Lu[2], Mai Sun[1] and Hongyu Lai[1]

[1]College of Engineering, Huazhong Agricultural University, Wuhan, China, [2]College of Resources and Environment, Huazhong Agricultural University, Wuhan, China

Introduction: Leaves are important organs for photosynthesis in plants, and the restriction of leaf growth is among the earliest visible effects under abiotic stress such as nutrient deficiency. Rapidly and accurately monitoring plant leaf area is of great importance in understanding plant growth status in modern agricultural production.

Method: In this paper, an image processing-based non-destructive monitoring device that includes an image acquisition device and image process deep learning net for acquiring *Brassica napus* (rapeseed) leaf area is proposed. A total of 1,080 rapeseed leaf image areas from five nutrient amendment treatments were continuously collected using the automatic leaf acquisition device and the commonly used area measurement methods (manual and stretching methods).

Results: The average error rate of the manual method is 12.12%, the average error rate of the stretching method is 5.63%, and the average error rate of the splint method is 0.65%. The accuracy of the automatic leaf acquisition device was improved by 11.47% and 4.98% compared with the manual and stretching methods, respectively, and had the advantages of speed and automation. Experiments on the effects of the manual method, stretching method, and splinting method on the growth of rapeseed are conducted, and the growth rate of rapeseed leaves under the stretching method treatment is considerably greater than that of the normal treatment rapeseed.

Discussion: The growth rate of leaves under the splinting method treatment was less than that of the normal rapeseed treatment. The mean intersection over union (mIoU) of the UNet-Attention model reached 90%, and the splint method had higher prediction accuracy with little influence on rapeseed.

KEYWORDS

leaf area, deep learning, image segmentation, monitor, effect of growth

1 Introduction

Leaves are very important organs in plants and are often monitored to reflect plant growth. In the late last century, leaf monitoring was mainly performed by manual measurement (Ross et al., 2000). With the development of computer technology, image processing is increasingly applied in the field of plant growth monitoring (Hamuda et al., 2016). Therefore, the accurate acquisition of leaf image information has become the focus of plant growth monitoring. Additionally, non-destructive continuous monitoring has become a hot topic in recent years. In particular, plant-specific tracking observation techniques are key to obtaining morphological parameters of plant dynamics over time (Ali et al., 2019; Su et al., 2020; Mohd Asaari et al., 2022). Monitoring crop growth is important for precision agriculture, allowing scientific and efficient cultivation to improve crop yields and reduce the waste of resources. Leaf area (LA) reflects the performance of mechanisms such as radiation shielding, water and energy exchange, plant growth, and biological productivity (Hsiao et al., 1970; Salah and Tardieu, 1997). Accurate continuous measurement of LA helps in understanding leaf ontogenesis, especially under multiple plant–environment interactions for researchers (Mielewczik et al., 2013; Friedli and Walter, 2015). Plants often encounter nutrient-improvised conditions, retarding leaf growth, decreasing LA, and reducing photosynthetic capacity. The reduction in LA usually arises first when plants are exposed to nutrient deficiency (Lu et al., 2016b). A reasonable estimation of LA and its variation would be useful to elucidate plant response mechanisms or phenotypes to nutrients or the environment, hence guiding the subsequent cultivation management.

The current methods for monitoring leaf growth are mainly manual destructive measurements and non-destructive monitoring based on computer vision. Traditional manual plant monitoring methods rely on the professional measuring experience of the people involved, and sophisticated instruments are used to measure various morphological parameters of the leaves (Jonckheere et al., 2004). It costs much time and money to train professionals with the ability to monitor plants, even well-trained workers will make mistakes in monitoring the plants' status, and there are few institutions dedicated to providing this type of training. In addition, the manual measurement of the measurement angle, leaf attitude, and other factors also affects the accuracy of the measurement. Another difficulty is the time required to manually assess the growth status of plants, which hinders rapid decision-making and large-scale assessments. Manual measurement often causes minimal damage to the leaf.

The continuous development of image processing technology in recent years and the appearance of many inexpensive image processing devices in the market allow the use of image processing technology in agriculture at a more extensive and easier stage. Non-destructive measurements based on computer vision include an online monitoring platform for plants based on 3D stereo reconstruction technology and a 2D image acquisition platform based on industrial cameras.

Plant monitoring methods based on 3D reconstruction techniques are used in many plant morphology studies. Apelt et al. (2015) built a 3D reconstruction platform using a light field camera to reconstruct the *Arabidopsis thaliana* wild-type in 3D and analyzed the information of individual leaf morphology and the time of leaf appearance in time series, providing a new idea for accurate plant morphology measurement and growth-related characteristics. Cuevas Velasquez et al. (2020) used a pair of stereo images obtained by the camera to reconstruct the rose in three dimensions, calculating the skeleton and branching structure as a basis for later pruning using the gardening robot. Although 3D reconstruction can reflect the overall morphology of the plant with high accuracy, there are problems, such as low efficiency and high cost for leaf measurement, due to expensive equipment and high setup and maintenance costs. Intensive monitoring in a very short amount of time is not possible because it takes considerable time to complete 3D reconstruction. Deep learning is a powerful tool for building 3D image processing models, but thus far, there are still problems such as too few samples and difficulties in annotating datasets when applying deep learning to 3D model reconstruction (Li et al., 2020).

Obtaining information from two-dimensional plant images through image processing algorithms is an efficient measurement method that uses digital images and time series to acquire plant phenotypic data using various methods and sensors in controlled environments and the field. There are precedents for using two-dimensional image processing techniques for LA, plant growth status, disease diagnosis, plant vigor, and postharvest vegetable quality monitoring (Pipatsitee et al., 2019; Baar et al., 2022). One of those methods is to use cameras to collect images and then import them into computer-side or smartphone-side software for batch processing, which further improves the accuracy of measurement and processing speed. Measurements are taken with a camera or smartphone, and the images are placed in software such as MATLAB or Image-Pro Plus to calculate the leaf area with the aid of a computer (Lu et al., 2016a; Badiger et al., 2022). Julian Schrader presented Leaf-IT, an application for measuring leaf area and other trait-related areas on smartphones (Schrader et al., 2017). Although this method can obtain high-precision leaf area data, it is destructive to the plant and cannot obtain continuously changing data of leaf area over time on the same plant. Another method is to fix light black beads around the leaf, use thin wire fixation to spread the leaf, and use a camera to photograph the leaf vertically (Mielewczik et al., 2013; Friedli and Walter, 2015). The stretching method mentioned above is usually used to measure the leaf area of soybean; in rapeseed, the results may be inaccurate. The force pull effect on leaf growth could not be assessed. Plant growth needs to be monitored as much as possible without affecting the natural growth state of the plant and at a high degree of automation. Therefore, there is some room for improvement in the stretching method in terms of the effect on the leaf.

In this study, a plywood-based leaf area acquisition device was designed to obtain leaf area based on a deep learning-based image semantic segmentation technique. The device adopts the form of clamping the whole leaf flat, reducing the influence of errors caused by leaf folds. Compared to the tensile method, the mechanical damage to the leaf is reduced. We use different deep learning models for the semantic segmentation of leaves and propose a U-

Net model that incorporates attention block, which has higher accuracy than other models. The accuracy of the splint measuring method is compared with that of the manual method and the stretching method. The effects of various methods on the growth of rapeseed were also investigated. The results showed that the splint method proposed in this paper had an acquisition error of leaf area within ±5%, which provides a feasible approach for non-destructive measurement of the rapeseed leaf area.

2 Materials and methods

2.1 Hardware setup

A complementary metal oxide semiconductor (CMOS) camera was used to obtain the leaf images and pass them into the leaf segmentation model located in the computer (Computer with AMD Ryzen 5 3500X processor, memory 16G, operating system Microsoft Windows 10 Professional, compiler Pycharm) for segmentation. Based on the switching power supply, LM2596 DC step-down module, electromagnetic relay and STM32 microcontroller, and other equipment, a set of night light control systems was built. The system is shown in Figure 1. The PC side and the microcontroller used serial communication, and the microcontroller obtained the light command, controlled the relay on and off to complete the light command response, and achieved an automatic night light switch.

2.2 Plant material

Rapeseed seedlings were grown in an environmentally controlled growth chamber located in the College of Resources and Environment, Huazhong Agricultural University, with a 14-h photoperiod under a photosynthetic photon flux density (PPFD) of 250 μmol m^{-2} s^{-1} at the leaf level. The temperature was 20°C during the day and 18°C at night, and the humidity in the greenhouse was controlled between 50% and 60%. Seeds were germinated for 7 days on gauze floating on the surface of deionized water in a dish. First, 80 uniform seedlings were transplanted into half-strength nutrient solution in 8.0-L black plastic containers. Seedlings were grown under a half-strength nutrient solution for 6 days before they were transplanted to a full-strength nutrient solution. The 13-day-old seedlings were transplanted to the full-strength nutrient solution, comprising, 3 mM of NH_4NO_3, 0.28 mM of Na_2HPO_4, 0.641 mM of NAH_2PO_4, 2 mM of KCl, 3.24 mM of $CaCl_2$, 2 mM of $MgCl_2$, 1 mM of Na_2SO_4, 4.6 μM of H_2BO_3, 9 μM of $MnCl_2$, 0.3 μM of $CuSO_4$, 0.8 μM of $ZnSO_4$, 0.1 μM of Na_2MoO_4, 0.1 μM of $Na_2MoO_4{\cdot}2H_2O$, 0.1 μM of $H_{32}Mo_7N_6O_{28}$, 0.05 mM of $FeSO_4$, and 0.05 mM of Na2EDTA. Five different treatments (CK\-N\-P\-K\-Mg) were used: control (CK, full nutrient), -N (0 mM of NH_4NO_3), -P (0 mM of Na_2HPO_4 and NAH_2PO_4), -K (0 mM of KCl), and -Mg (0 mM of $MgCl_2$). The nutrient solution was renewed every 3 days. During hydroponic cultivation, the nutrient solution was aerated for 0.5 h with an air pump that bubbled air through air stones every 2 h to maintain a water oxygen content of approximately 8.0 mg/L.

2.3 Measuring methods

Three images were collected for each plant by selecting a leaf, and the daytime and night-time acquisitions are listed in Table 1. In the stretching method, leaves were stretched by five small clips with a proper counterweight, and images were captured by a camera installed directly above the leaves.

TABLE 1 Acquisition of image datasets.

Day/night group	Method			
	Manual	Splint	Stretching	Total
-CK	36	36	36	108
-N	36	36	36	108
-P	36	36	36	108
-K	36	36	36	108
-MG	36	36	36	108
Total	180	180	180	540

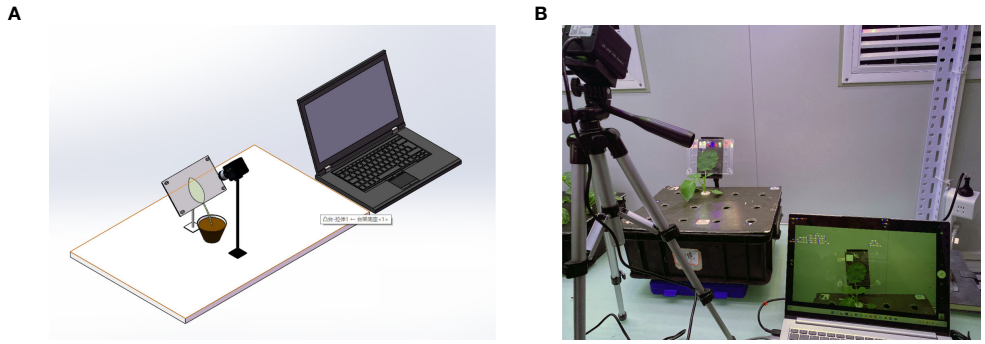


FIGURE 1
(A) 3D view of the device. **(B)** Image acquisition device. This device consists of an industrial camera, a splint, and a laptop. The computer drives the industrial camera, and the image of the blade is captured by the image sensor in the camera and transferred to the laptop for storage through the USB connection cable.

2.3.1 Manual method and stretching method

In the manual method, the front view image of the rapeseed leaf was taken by a mobile phone and put into computer software for analysis shown in Figure 2B. First, the preprepared calibration board was placed behind the rapeseed leaves to be tested, and a smartphone was used to take a frontal image of the rapeseed, which was stored in the phone's built-in memory. After photos were taken, the images in the phone were exported to computer software, IPP6.0 (Image-pro Plus 4.5 software) (Media Cybernetics, Silver Spring, MD, USA) to process the image. In the software, first, the leaf area was manually selected, the number of pixels PNL_C (pixel number of a leaf) was written down, and then the area of the known area S_C (5 mm × 5 mm) was selected. The number of pixels is denoted as PNC_C (pixel number of calibration objects) and then the leaf area as LA_C:

$$LA_c = \frac{PNL_c \times S_c}{PNC_c}. \quad (1)$$

LA_C is the leaf area, PNL_C is the pixel number of leaf, and PNC_C is the pixel number of calibration objects.

In the stretching method which is shown in Figure 2C, each leaf was fixed to the camera's focusing plane, supported by a plate with a white background, and evenly clamped to the five points of the leaf by five small clips connected by thin lines with a counterweight consisting of a centrifugal tube filled with an appropriate amount of water, hanging around the edge of the ring. The camera was fixed 30 cm above the leaf, and the lens was parallel to the leaf. A set of computer-controlled lighting systems was set up on the bench to enable the night camera to work properly. Before an image was obtained, the calibration board was fixed on the background plate, a 5 mm × 5 mm area on the calibration board was taken as the calibration area (S_{St}), and the number of pixels PNS_{St} (pixel number of stretching) were recorded. The actual area of the marked area is S_{St}. The number of pixels in the leaf area calculated in the algorithm was taken as PNL_{St} (pixel number of stretching methodPixel number of Stretching Method) and then the leaf area as LA_{St}:

$$LA_{St} = \frac{PNL_{St} \times S_{St}}{PNC_{St}}. \quad (2)$$

LA_{St} is the leaf area of the stretching method, and PNL_{St} is the pixel number of the leaf area in the stretching method.

2.3.2 Splint method

In this method, the monitored leaf was fixed to the camera imaging plane by two transparent acrylic plates, which were fixed to the greenhouse stand by a rubber base, and the two acrylic plates were connected by bolts and nuts it is shown in Figure 2A. Additionally, the plywood lamination gap can be adjusted by rotating the bolts to suit different thicknesses of the leaf. During the image acquisition process, the light was fixed by a variable bracket approximately 50 cm directly above the camera and approximately 100 cm away from the object. The device was set to leave the light on during the day, turn on the light 1 s before photographing the image at night, and turn off the light 1 s after images were taken. Before image capture, the calibration board was fixed on the acrylic plate, a picture was taken as the calibration image, and the 5 mm × 5 mm size area was employed on the calibration board as the calibration area (S_{Sp}). We note down the number of pixel points was taken as PN_{Sp} (pixel number of the splint method) and the number of pixel points as PNL_{Sp} (pixel number of the leaf area) as calculated in the algorithm. Then, the leaf area was calculated as follows:

$$LA_{Sp} = \frac{PNL_{Sp} \times S_{Sp}}{PNC_{Sp}}. \quad (3)$$

LA_{Sp} is the leaf area of the splint method, and PNL_{Sp} is the pixel number of the leaf area in the splint method.

2.3.3 Calibration leaf area

To compare the effectiveness of the three measurement methods, the measured leaves were cut out and placed into the scanner for scanning, and leaf area S was measured using the leaf analysis software as the standard value. The measurement accuracy (Acc) is expressed as follows:

$$Acc = \frac{|S - S_t|}{S} \times 100\,\%. \quad (4)$$

2.4 Image segmentation model

Image annotation is an important step in the extraction of rapeseed leaves using deep learning methods. The semantic segmentation of the image is compiled by using the Labelme software, which separates the rapeseed leaves from the image background and obtains the json format file with the boundary points of the leaves, and the masked image of the rapeseed leaves is obtained after the segmentation process. The rapeseed leaf segmentation dataset was randomly divided into a training set, test set, and validation set with a ratio of 6:2:2. The training set, test set, and validation set contain 648, 216, and 216 images, respectively, for a total of 1,080 images.

2.4.1 PSPNet

PSPNet was proposed in 2017 by Zhao et al. (2017) and was improved based on fully convolutional network (FCN). The backbone part uses ResNet50 as the feature extraction model, and the feature layers extracted from the backbone part are fused under four different scales using a pyramid module. The pyramid module is actually a four-level module, the top level is the global average pooling, and levels 2, 3, and 4 divide the input feature layers into 2×2, 3×3, and 6×6 subregions, respectively. The pooling was averaged for each subregion separately. Finally, a 3×3 convolution was used for feature integration, and a 1×1 convolution was used for channel adjustment. The PSPNet model structure is shown in Figure 3.

10.3389/fpls.2023.1163700

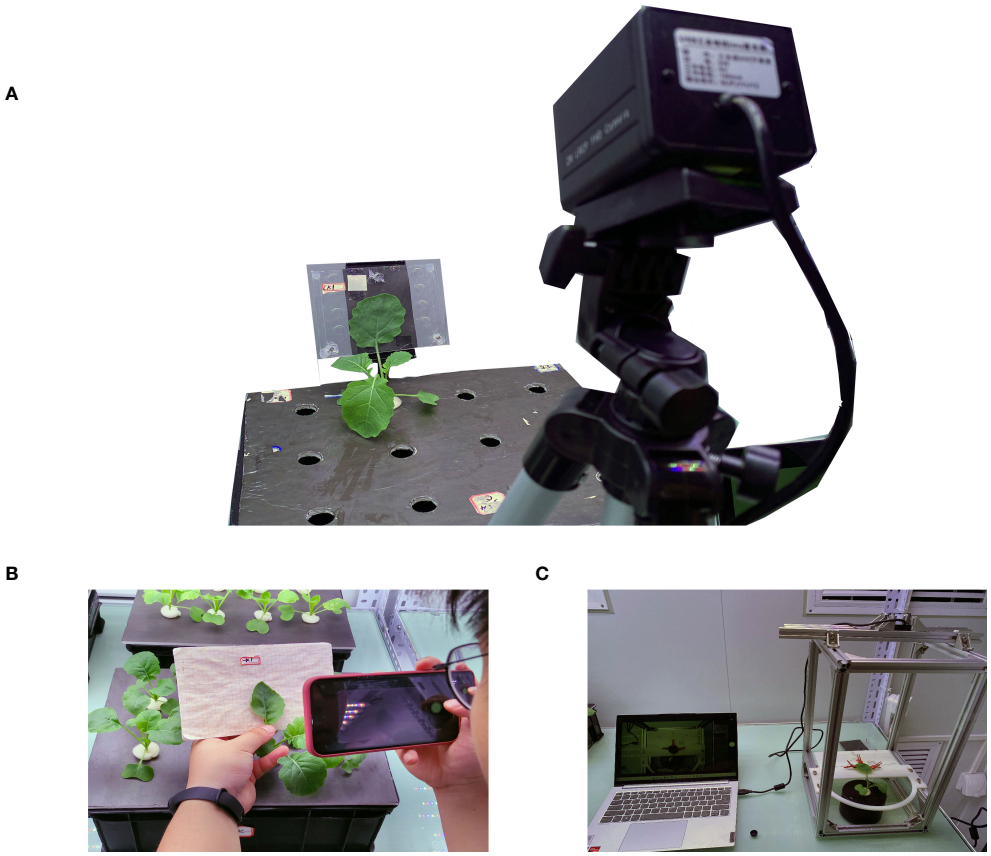


FIGURE 2
(A) Splint method. **(B)** Manual method. **(C)** Stretching method.

2.4.2 DeepLab v3+

DeepLab v3+ is a new semantic segmentation algorithm that was introduced by Google in 2018, improving on the DeepLab v1-3. Cai introduced the attention mechanism into the DeepLab v3+ model as shown in Figure 4, enhanced the feature information of strawberry images, and improved the segmentation accuracy by adjusting the weights of the feature channels during the propagation of the neural model through the attention mechanism (Cai et al., 2022). Peng et al. applied DeepLab v3+ to the segmentation of litchi branches and used migration learning and data enhancement methods to accelerate model convergence and improve model robustness (Peng et al., 2020).

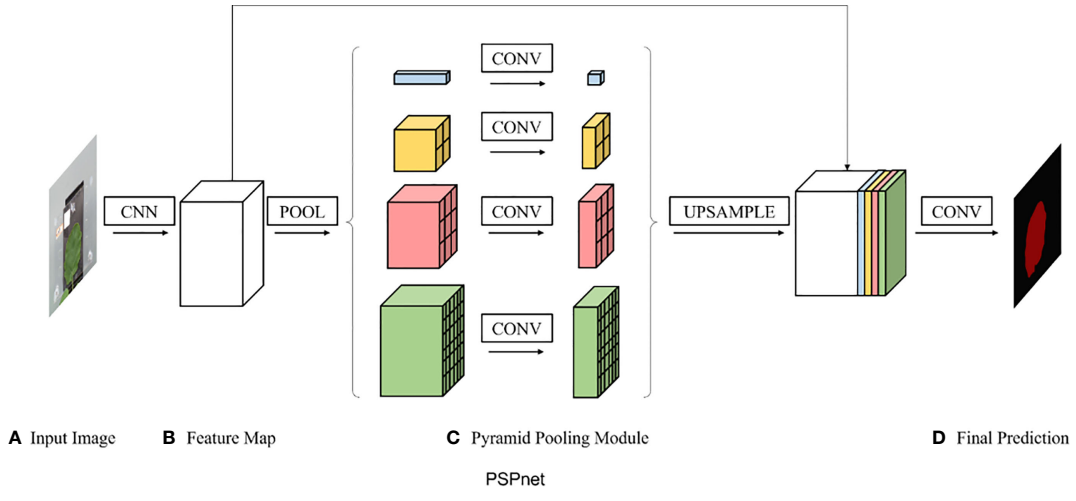


FIGURE 3
PSPNet structure.

10.3389/fpls.2023.1163700

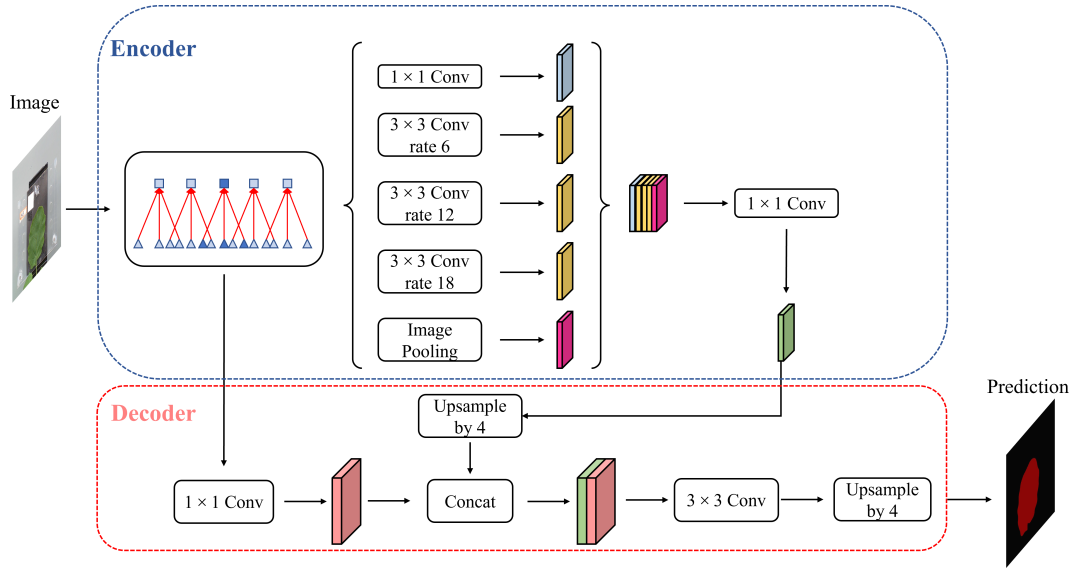


FIGURE 4
DeepLab V3+ structure.

Therefore, the DeepLab v3+ model was selected as a semantic segmentation model method in this study.

2.4.3 U-Net

For the segmentation model, we used U-Net architecture, which is proposed by Ronneberger et al. (2015), a deep learning-based image segmentation model. Deep learning-based semantic segmentation algorithms classify images at the pixel level. The input image is processed by the deep neural model. The pixels in the image are encoded by the convolution layer and pooling layer in the downsampling process of the model and then decoded by deconvolution in the upsampling process. Finally, the segmented image is obtained.

As shown in Figure 5, the U-Net is a deep learning model with a typical encoding–decoding structure. It has a left–right symmetric structure, similar to the letter, with the encoding part of the model on the left and the decoding part on the right. The advantage of the U-Net model is that it can achieve high segmentation accuracy with relatively less data.

The "encoding" part uses VGG16 as the main feature extraction part, with five layers. Each layer uses two 3 * 3 convolutional kernels, each followed by a rectified linear unit (ReLU) activation function and a 2 * 2 maximum pooling operation. The "decoding" part uses the five effective feature layers obtained from the "encoding" part for feature fusion by upsampling and stacking the feature layers.

The encoding part consists of four submodules, each of which downsamples the model of the previous level by a factor of two, and the resolution of the image decreases to one-half of the original with every module. The decoding part is similar in structure to the encoding part and consists of four submodules, each of which

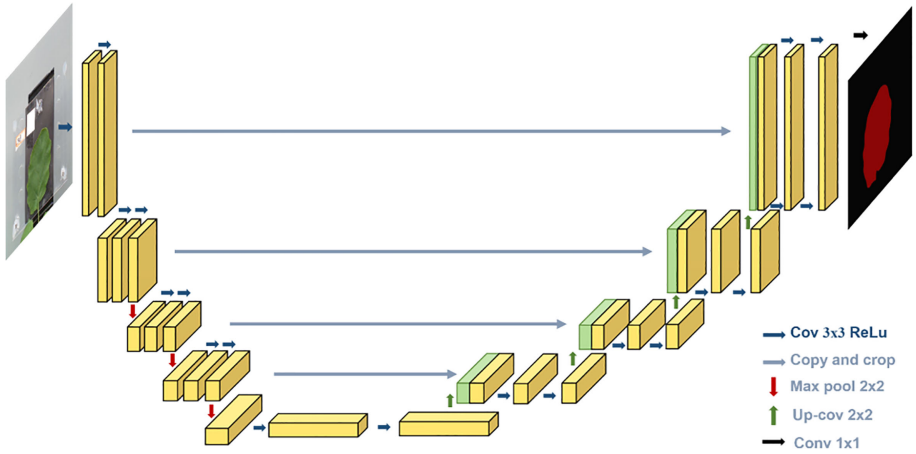


FIGURE 5
U-Net structure.

upsamples the input image by a factor of two, and with each passing submodule, the resolution of the image rises to two times that of the input image. The loss function with boundary weights in the loss function is formulated as follows:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}, \tag{5}$$

where A is the *a priori* mask and B is the predicted mask.

2.4.4 Attention block

The attention model was first introduced in the seq2seq model (Sutskever, et al., 2014). It is now widely used in different types of machine learning tasks such as natural language processing and image processing as well as speech recognition. According to the different domains where the attention mechanism is applied, the attention weights are applied in different ways and locations, and the attention mechanism is divided into three kinds: spatial domain, channel domain, and hybrid domain. Among them, channel attention has a strong generalization in the image processing domain, so this paper adds channel attention to the channel domain. Channel attention is similar to applying a weight to each channel's feature map, which represents the relevance of that channel to the key information, and the larger the weight, the higher the relevance. By adding the SE attention block in the fusion stage of the shallow features and deep features in the U-Net model, the shallow features generated in the downsampling procedure are processed by the channel attention and then fused with the subsequent features to achieve the goal of improving the prediction accuracy, and the U-Net-attention structure is shown in Figure 6A.

After the feature map is passed in, it is changed into a C-channel feature map with the length and width of H and W, respectively, by one convolution and then into a 3D matrix of size 1◊1◊C by one global average pooling. Then, the weights of each channel are fixed to between 0 and 1 by two full joins and one sigmoid, at which time the matrix corresponds to the weights of each channel in the input layer, and finally, these weights are multiplied by the original input feature layer, which completes channel attention processing. The SE block is shown in Figure 6B.

2.5 Evaluation of the model

In this paper, we use mean intersection over union (mIoU), mPA, mPrecision, and recall, which are usually used for segmentation tasks as the evaluation metrics, and each metric is calculated as follows:

$$\text{MIoU} = \frac{\text{TP}}{\text{FN+TP+FP}}, \tag{6}$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \times 100\,\%, \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+TN}} \times 100\,\%, \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 100\,\%. \tag{9}$$

TP, *FP*, *FP*, and *FN* are from the confusion matrix. The confusion matrix is widely used in the field of machine learning,

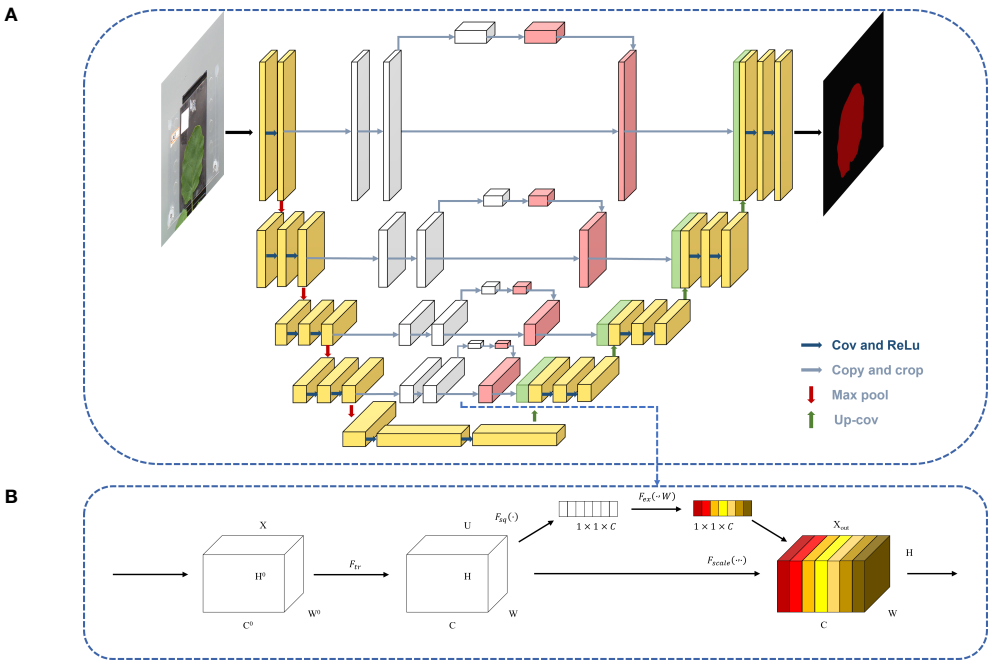


FIGURE 6
The U-Net-Attention model.

it is also known as a likelihood matrix or error matrix, and it is a visualization tool.

2.6 Effect of the measuring method on the growth experiment

To investigate the effect of continuous monitoring on rapeseed leaves, we also conducted a 7-day continuous monitoring experiment in which leaf area information was collected daily, and the change in the area for seven consecutive days was used as an indicator to evaluate the effect of the monitoring device on rapeseed growth.

We selected the same batch of rape seeds for cultivation, from which 12 seedlings with similar growth were selected and transferred into specific cultivation containers when they were 8 days old. When the seedlings reached 14 days old, they were treated in batches and divided into blank, stretch, and splint treatments, with the following nutrient solution ratios.

Monitoring started at the seedling age of 24 days and continued for eight consecutive days. The leaf area was measured every morning at 08:00. The measurement method was the manual method.

3 Results

3.1 Accuracy of different measuring methods

The result showed that the median error rate of all three kinds of measuring methods was within 10%. The error rates of the three methods differed substantially, where a negative error rate indicates that the area measured by this method was less than the standard area, and a positive error rate indicates that the area measured by this method was greater than the standard value. The results showed that the stretching method had a negative error for leaf area measurement. The measured leaf area data were smaller than the area data obtained from the analysis with the leaf area scanning software. As shown in Figure 7A, in the "ck" group, during the day, the average error of the manual method was 9.13%, the average error of the stretching method was 4.60%, and the average error of the splint method measurement was 1.25%. The same situation can be seen at night. During the night, the average error of the manual method was 15.10%, the average error of the stretching method was 6.66%, and the average error of the splint method measurement was 0.05%. In other words, the average error rate of the manual method was 12.12%, the average error rate of the stretching method was 5.63%, and the average error rate of the splint method was 0.65%.

The error rate of the manual method was positive in all five treatments. Another interesting aspect of this graph was that, when using the stretching method, the predicted area was always lower than the area predicted by the scanning method. Therefore, when images of leaves were obtained, the clips would cover part of the leaf, and during the image process, this part of the leaf would not be classified to be the leaf part. Then, the area of the leaf predicted by the stretching method was lower than the area measured by the scanner. What is striking in this figure is that the error rate of the SP method was ±%5, which is much better than that of either of the other two methods.

Another point that needed to be considered was that the error rate varied with the treatment that the plants were taken in, and the explanation for this phenomenon is that the leaves' shape and size vary when their images are captured. This is particularly true when they are been under nutrient stress such as a lack of nitrogen, phosphorus, potassium, and magnesium. The young plant of rapeseed shows different symptoms when deficient in nutrients, especially on the leaves. When they lack nitrogen, the leaves are yellow; when they lack potassium, the marginal part of the leaves is yellow; when they lack phosphorus, the leaves are curled; and when they lack magnesium, the leaves are partly yellow. Therefore, leaf deficiency symptoms have some influence on the prediction of leaf area, and there is a normal phenomenon in which the error rate is different when the treatment varies.

According to the results mentioned above, the area measured by the SP method is closer to the real area of the leaf.

3.2 Model training and validation

To make a comparison with the U-Net-attention algorithm proposed in this study, we also used deep learning segmentation methods such as PSPNet to process the images. By comparing the model segmentation indices, we finally obtained the segmentation method with higher accuracy. According to Table 2, we can see that the proposed U-Net-attention semantic segmentation model with a fused attention mechanism had more advantages.

The segmentation result is shown in Figure 8. The semantic segmentation model based on the U-Net-attention model well-segmented the leaf area in the image, and the application effect in the splint method was suitable, which could meet the requirement of calculating the area accuracy. HRnet could segment the leaf region roughly, but there was some loss of leaf edge detail. DeepLab v3+ had a slightly worse segmentation effect than HRnet, and the degree of edge loss was more serious. PSPNet has a better segmentation effect in the stretching and manual segmentation.

The models used in the above three segmentation models did not perform well in the dataset of this experiment. The U-Net model performed well on the dataset with a small sample size. As shown in Figure 7, the model segmentation effect of the U-Net and U-Net-attention model was significantly better than the other three models, but the U-Net model without the attention mechanism appears less vacant for rapeseed segmentation, and the addition of the attention mechanism eliminates this phenomenon.

We also tested the time to process the images for each prediction model. There is little difference in the running time of U-Net attention with the addition of the attention mechanism compared to that without it. Although HRnet, DeepLab v3+, and PSPNet have improved versions in prediction time due to U-Net and its addition of the attention mechanism since the current study focuses on leaf segmentation, the segmentation accuracy is used as

10.3389/fpls.2023.1163700

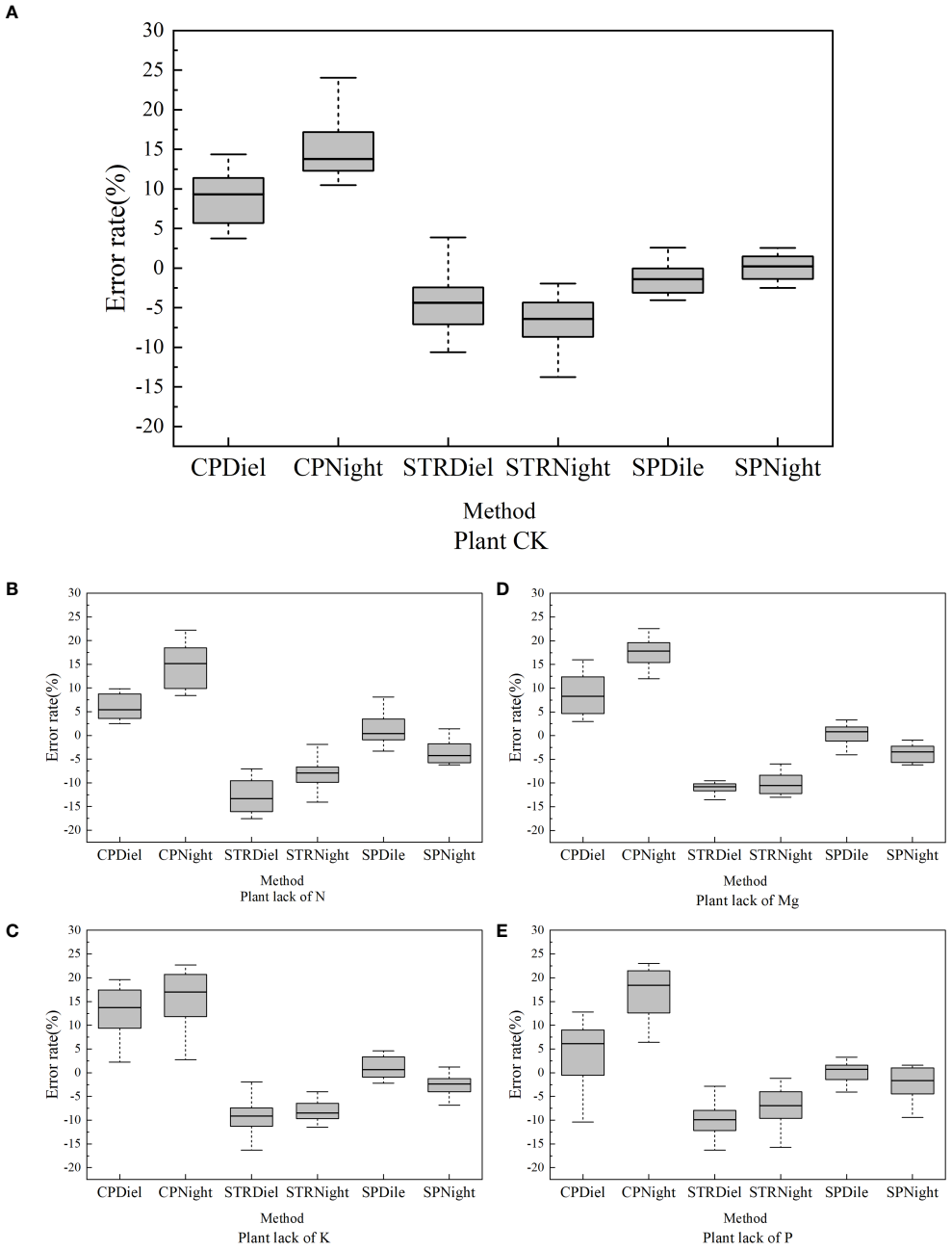


FIGURE 7
Box diagram of image segmentation error rate of rapeseed under different nutritional conditions. **(A)** Error rate under normal nutritional conditions. **(B)** Error rate under nitrogen deficiency. **(C)** error rate under potassium deficiency. **(D)** Error rate under magnesium deficiency. **(E)** Error rate under phosphorus deficiency.

the main evaluation index, and U-Net-attention model is used as the optimal segmentation model.

3.3 Effects of different measuring methods on growth

The purpose of this experiment was to compare the effect on the growth rate of plants with different measuring methods on oilseed rapeseed. We started monitoring the leaf area on the first day as a standard and compared the change in leaf area for the following 7 days.

The growth rate of rapeseed was measured under three methods, with two replicates of each method, and the middle value of three repetitions was taken for analysis. The leaf area of each leaf was measured by various measuring methods at 16:00 every day. Figure 9 provides an overview of the effects of three measuring methods on the growth rate of rapeseed. Figure 9 reveals that there was a substantial difference between the manual method, stretching method, and splint method. As seen in the figure, there

TABLE 2 Segmentation results on different models.

Model	mIoU	mPA	mPrecision	Recall	Time (s)
HRnet	82.37%	84.31%	94.45%	68.78%	1.09
DeepLab v3+	88.50%	90.21%	95.99%	80.55%	0.76
PSPNet	75.88%	77.42%	95.70%	54.93%	1.08
U-Net	98.09%	99.32%	97.53%	98.74%	1.39
U-Net-Attention	98.58%	99.42%	98.34%	98.90%	1.37

was a steady rise in all three plant treatments, which means that there was no apparent inhibition of plant growth regardless of the measuring method we used.

What is striking is that compared with the manual method, plants measured by the splint method had a lower growth rate than plants measured by the manual method and the stretching method. This may be due to the effect that the splint had on the leaf in vertical growth since the splint applies restraint in the vertical direction of the leaf to ensure that the leaf was flat. Although this had some effect on the growth of rapeseed leaf, it ensured the accuracy of leaf area measurement.

4 Discussion

Rapid and non-destructive measurement of leaf area is essential for monitoring plant growth rate, and growth rate monitoring provides agricultural producers with a means to monitor plant status and growth so they can more accurately plan and manage the crop production process (Karimi and Siddique, 1991).

In this study, an image processing technology-based rapeseed leaf area measurement method was proposed. The U-Net-Attention learning model with an attention block was used for an initial segmentation of the original image; then, an image processing algorithm was used to perform binarization and hole-filling operations on the image; finally, the leaf area based on the number of leaf pixels was compared with the calibrated pixels. Based on this measurement method, leaf area measurement and continuous area change monitoring of rapeseed under different measurement conditions were completed. With the leaf area obtained from the scanner as the benchmark, the average accuracy of the leaf segmentation of the algorithm proposed in this paper was 96.77%, which was higher than the accuracy of other segmentation models. Meanwhile, rapeseed leaf monitoring experiments based on the splint method obtained more objective patterns, which laid the foundation for further leaf growth monitoring experiments based on image processing.

Other than that, the manual method (Wang et al., 2019; Lu et al., 2022) and stretching method (Ainsworth et al., 2005; Walter et al., 2008) had been used in the previous research. Manual measurement methods require cutting the leaves from the plant or using a camera to photograph the leaves in the original location; then, using professional image processing software can accurately measure the area of the cut leaves, but the absence of the leaves destroyed the natural growth state of the plant, so continuous observation in its natural state is not possible. Although the use of a camera to photograph the leaves in the original location ensures that the natural growth of the plant state maximally, the operator cannot make sure that the camera is perpendicular to the leaf and

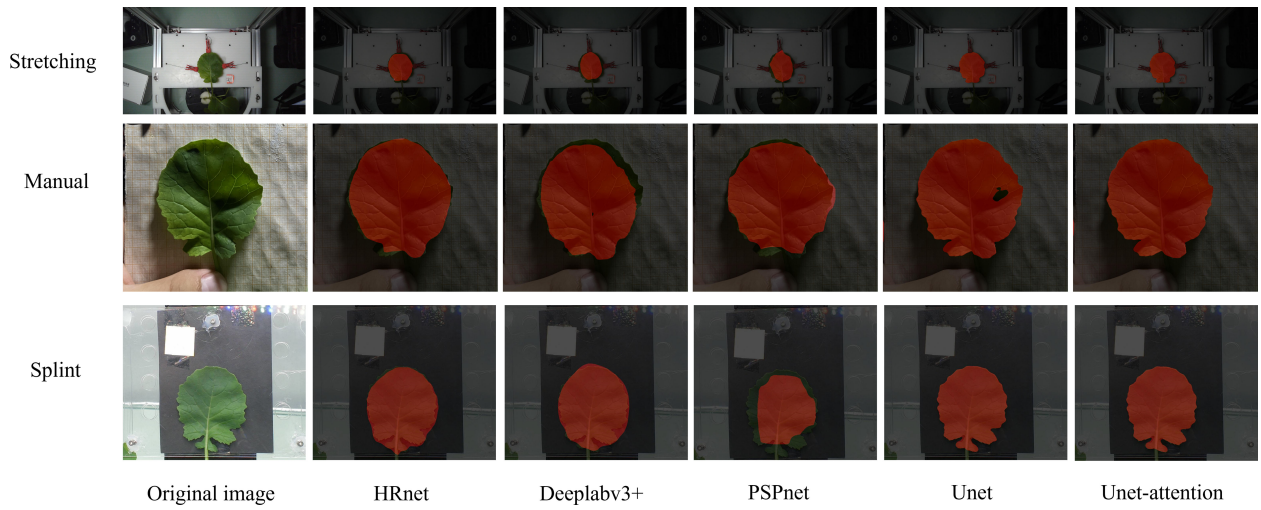


FIGURE 8
Image segmentation results.

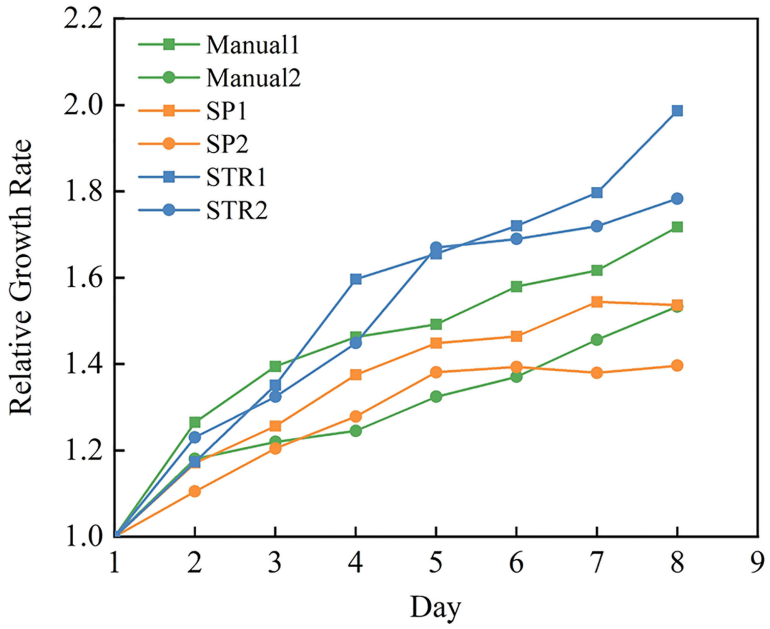


FIGURE 9
Growth rate of the three measuring methods.

the leaf for the flat state. In Figure 9, we can see that the area measured by the manual method was larger than the ground true area. The stretching method could ensure the flatness of the leaf and camera when taking images, but the tension of the leaf caused a slight increase in leaf growth rate according to Figure 9. The splint method proposed in this paper can accomplish the measurement of rapeseed leaf area at a lower cost and reduce labor intensity. At the same time, it had less effect on leaf growth compared to the stretching method and manual method. Although the splint method had a slight effect on rapeseed leaf growth, according to the measurement accuracy, the splint method was superior to the other two measuring methods.

Many excellent methods have been proposed in the research for leaf monitoring. These include leaf area measurement by stretching the leaves (Ainsworth et al., 2005; Walter et al., 2008) and obtaining plant morphological information using 3D modeling techniques (Apelt et al., 2015; An et al., 2017; Boukhana et al., 2022). As can be seen from Figure 9, the stretching measurement for the leaf may destroy the original growth pattern of the leaf, while the leaf growth is a continuous process, and the rate of area change is a measure of the growth state, which requires the measurement of leaf area without affecting the original growth state of the leaf as much as possible. 3D modeling technologies can reconstruct the leaf morphology and obtain the leaf area without touching the plant, and the impact on the plant caused by its impact on the plant is almost negligible. However, the cost of expensive equipment makes it difficult to deploy this method on a large scale in the laboratory. Deep learning techniques, which have gradually become a hot research direction in recent years, can solve most of the problems that are difficult to overcome in traditional image processing methods (Baar et al., 2022; Sapoukhina et al., 2022; Tamvakis et al., 2022).

Many deep learning models had been used in the monitoring task of plants (Halstead et al., 2021; Sapoukhina et al., 2022). Some common models such as U-Net, DeppLab v3+, PSPNet, and HRnet had been widely used in leaf segmentation and disease diagnosis tasks. Deep learning model accuracy comparison in this paper showed that the U-Net model had the best performance, while PSPNet had the worst results. In addition, U-Net and DeepLab v3+ models perform well in the leaf disease segmentation task (Divyanth et al., 2023). Therefore, U-Net and DeepLab v3+ could be the preferred models for plant leaf image segmentation tasks. In this paper, the attention mechanism was incorporated into the U-Net model to form the U-Net-Attention model, and it can be seen from Table 2 that the U-Net-Attention model had higher accuracy. Incorporating attention mechanisms into models to improve segmentation accuracy is a more common approach in deep learning, and in order to adapt to specific segmentation tasks, Mishra added the attention mechanism into the original model, which further improves the segmentation accuracy (Mishra et al., 2021). In future work, we can try to add the attention module to different positions in the model to obtain better segmentation effects.

Figure 9 shows that although the splinting method proposed in this paper can increase the measurement accuracy of rapeseed leaf to an extent, there is still a certain inhibition effect of splinting on leaf growth compared to the control group. In future work, we need to further explore better ways to fix the leaves and make non-destructive measurements with minimum effect on leaf growth.

The use of non-contact methods to obtain plant phenotype information has always been a concern for scholars, and non-contact measurement methods are a key aspect of the non-destructive measurement of plant phenotypes. With the development of image processing technology, an increasing

number of high-throughput computer vision-based plant monitoring devices are appearing in the agricultural field, which enables researchers to reach a new level of research on plant growth process monitoring(Zhang et al., 2021).

However, the most advanced methods currently focus on a two-dimensional analysis and three-dimensional model building of plant structure and morphology. There are fewer studies on the growth monitoring of individual plant organs. The leaves, stalks, flowers, and fruits of plants show different behaviors based on environmental stresses, which affect the physiological and biochemical processes of the whole plant and influence the growth of the plant. The splint method performed in this paper provided a good way to monitor the phenotypic characteristics of plants. Therefore, in the future, this method could be used in plant disease monitoring.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ML, YL, and ZL conceived the idea, proposed the method, and revised the paper. MS and HL contributed to the preparation of the experiment and the acquisition of data. ML wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the National Natural Science Foundation of China (32272820) and the National oilseed rape industry technology system special (CARS-12).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ainsworth, E. A., Walter, A., and Schurr, U. (2005). Glycine max leaflets lack a base-tip gradient in growth rate. *J. Plant Res.* 118, 343–346. doi: 10.1007/s10265-005-0227-1

Ali, M. M., Bachik, N. A., Muhadi, N. A., Tuan Yusof, T. N., and Gomes, C. (2019). Non-destructive techniques of detecting plant diseases: a review. *Physiol. Mol. Plant Pathol.* 108, 101426. doi: 10.1016/j.pmpp.2019.101426

An, N., Welch, S. M., Markelz, R. J. C., Baker, R. L., Palmer, C. M., Ta, J., et al. (2017). Quantifying time-series of leaf morphology using 2d and 3d photogrammetry methods for high-throughput plant phenotyping. *Comput. Electron. Agric.* 135, 222–232. doi: 10.1016/j.compag.2017.02.001

Apelt, F., Breuer, D., Nikoloski, Z., Stitt, M., and Kragler, F. (2015). Phytotyping4d: a light-field imaging system for non-invasive and accurate monitoring of spatio-temporal plant growth. *Plant J.* 82, 693–706. doi: 10.1111/tpj.12833

Baar, S., Kobayashi, Y., Horie, T., Sato, K., Suto, H., Watanabe, S., et al. (2022). Non-destructive leaf area index estimation *via* guided optical imaging for large scale greenhouse environments. *Comput. Electron. Agric.* 197, 106911. doi: 10.1016/j.compag.2022.106911

Badiger, M., Kumara, V., Shetty, S. C. N., and Poojary, S. (2022). Leaf and skin disease detection using image processing. *Global Transit. Proc.* 3, 272–278. doi: 10.1016/j.gltp.2022.03.010

Boukhana, M., Ravaglia, J., Hétroy-Wheeler, F., and De Solan, B. (2022). Geometric models for plant leaf area estimation from 3d point clouds: a comparative study. *Graphics Visual Comput.* 7, 200057. doi: 10.1016/j.gvc.2022.200057

Cai, C., Tan, J., Zhang, P., Ye, Y., and Zhang, J. (2022). Determining strawberries' varying maturity levels by utilizing image segmentation methods of improved deeplabv3+. *Agronomy* 12, 1875. doi: 10.3390/agronomy12081875

Cuevas Velasquez, H., Gallego Sánchez, A. J., and Fisher, B. (2020). Segmentation and 3D reconstruction of rose plants from stereoscopic images. *Comput. Electron. Agric.* 171, 105296. doi: 10.1016/j.compag.2020.105296

Divyanth, L. G., Ahmad, A., and Saraswat, D. (2023). A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery. *Smart Agric. Technol.* 3, 100108. doi: 10.1016/j.atech.2022.100108

Friedli, M., and Walter, A. (2015). Diel growth patterns of young soybean (glycine max) leaflets are synchronous throughout different positions on a plant. *Plant Cell Environ.* 38, 514–524. doi: 10.1111/pce.12407

Halstead, M., Ahmadi, A., Smitt, C., Schmittmann, O., and McCool, C. (2021). Crop agnostic monitoring driven by deep learning. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.786702

Hamuda, E., Glavin, M., and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* 125, 184–199. doi: 10.1016/j.compag.2016.04.024

Hsiao, T. C., Acevedo, E., and Henderson, D. W. (1970). Maize leaf elongation: continuous measurements and close dependence on plant water status. *Science* 168, 590–591. doi: 10.1126/science.168.3931.590

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., et al. (2004). Review of methods for *in situ* leaf area index determination. *Agric. For. Meteorol.* 121, 19–35. doi: 10.1016/j.agrformet.2003.08.027

Karimi, M. M., and Siddique, K. (1991). Crop growth and relative growth rates of old and modern wheat cultivars. *Aust. J. Agric. Res.* 42, 13. doi: 10.1071/AR9910013

Li, Z., Guo, R., Li, M., Chen, Y., and Li, G. (2020). A review of computer vision technologies for plant phenotyping. *Comput. Electron. Agric.* 176, 105672. doi: 10.1016/j.compag.2020.105672

Lu, Z., Hu, W., Ye, X., Lu, J., Gu, H., Li, X., et al. (2022). Potassium regulates diel leaf growth ofbrassica napus by coordinating the rhythmic carbon supply and water balance. *J. Exp. Bot.* 73, 3686–3698. doi: 10.1093/jxb/erac060

Lu, Z., Lu, J., Pan, Y., Li, X., Cong, R., Ren, T., et al. (2016a). Genotypic variation in photosynthetic limitation responses to k deficiency of brassica napus is associated with potassium utilisation efficiency. *Funct. Plant Biol.* 43, 880. doi: 10.1071/FP16098

Lu, Z., Lu, J., Pan, Y., Lu, P., Li, X., Cong, R., et al. (2016b). Anatomical variation of mesophyll conductance under potassium deficiency has a vital role in determining leaf photosynthesis. *Plant Cell Environ.* 39, 2428–2439. doi: 10.1111/pce.12795

Mielewczik, M., Friedli, M., Kirchgessner, N., and Walter, A. (2013). Diel leaf growth of soybean: a novel method to analyze two-dimensional leaf expansion in high temporal resolution based on a marker tracking approach (martrack leaf). *Plant Methods* 9, 30. doi: 10.1186/1746-4811-9-30

Mishra, P., Sadeh, R., Bino, E., Polder, G., Boer, M. P., Rutledge, D. N., et al. (2021). Complementary chemometrics and deep learning for semantic segmentation of tall and wide visible and near-infrared spectral images of plants. *Comput. Electron. Agric.* 186, 106226. doi: 10.1016/j.compag.2021.106226

Mohd Asaari, M. S., Mertens, S., Verbraeken, L., Dhondt, S., Inzé, D., Bikram, K., et al. (2022). Non-destructive analysis of plant physiological traits using hyperspectral imaging: a case study on drought stress. *Comput. Electron. Agric.* 195, 106806. doi: 10.1016/j.compag.2022.106806

Peng, H., Xue, C., Shao, Y., Chen, K., Xiong, J., et al. (2020). Semantic segmentation of litchi branches using deeplabv3+ model. *IEEE Access* 8, 164546–164555. doi: 10.1109/ACCESS.2020.3021739

Pipatsitee, P., Eiumnoh, A., Praseartkul, P., Ponganan, N., Taota, K., Kongpugdee, S., et al. (2019). Non-destructive leaf area estimation model for overall growth performances in relation to yield attributes of cassava (manihot esculenta cranz) under water deficit conditions. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 47, 580-591. doi: 10.15835/nbha47311487

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentationMedical image computing and computer-assisted intervention – MICCAI 2015. MICCAI 2015. . *Lecture Notes Comput. Sci.* (Cham) 9351. doi: 10.1007/978-3-319-24574-4_28

Ross, J., Ross, V., and Koppel, A. (2000). Estimation of leaf area and its vertical distribution during growth period. *Agric. For. Meteorol.* 101, 237–246. doi: 10.1016/S0168-1923(00)00089-7

Salah, H., and Tardieu, F. (1997). Control of leaf expansion rate of droughted maize plants under fluctuating evaporative demand - a superposition of hydraulic and chemical messages? *Plant Physiol.* 114, 893–900. doi: 10.1104/pp.114.3.893

Sapoukhina, N., Boureau, T., and Rousseau, D. (2022). Plant disease symptom segmentation in chlorophyll fluorescence imaging with a synthetic dataset. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.969205

Schrader, J., Pillar, G., and Kreft, H. (2017). Leaf-it: an android application for measuring leaf area. *Ecol. Evol.* 7, 9731–9738. doi: 10.1002/ece3.3485

Su, W., Slaughter, D. C., and Fennimore, S. A. (2020). Non-destructive evaluation of photostability of crop signaling compounds and dose effects on celery vigor for precision plant identification using computer vision. *Comput. Electron. Agric.* 168, 105155. doi: 10.1016/j.compag.2019.105155

Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to Sequence Learning with Neural Networks. arXiv:1409.3215

Tamvakis, P. N., Kiourt, C., Solomou, A. D., Ioannakis, G., and Tsirliganis, N. C. (2022). Semantic image segmentation with deep learning for vine leaf phenotyping. *Ifac-Papersonline* 55, 83–88. doi: 10.1016/j.ifacol.2022.11.119

Walter, A., Christ, M. M., Rascher, U., Schurr, U., and Osmond, B. (2008). Diel leaf growth cycles in clusia spp. are related to changes between c3 photosynthesis and crassulacean acid metabolism during development and during water stress. *Plant Cell Environ.* 31, 484-491. doi: 10.1111/j.1365-3040.2008.01777.x

Wang, Y., Jin, G., Shi, B., and Liu, Z. (2019). Empirical models for measuring the leaf area and leaf mass across growing periods in broadleaf species with two life histories. *Ecol. Indic.* 102, 289–301. doi: 10.1016/j.ecolind.2019.02.041

Zhang, R., Li, G., Wunderlich, T., and Wang, L. (2021). A survey on deep learning-based precise boundary recovery of semantic segmentation for images and point clouds. *Int. J. Appl. Earth Observation Geoinformat.* 102, 102411. doi: 10.1016/j.jag.2021.102411

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. yramid Scene Parsing Network. arXiv:1612.01105

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 13 October 2023
DOI 10.3389/fpls.2023.1219476

Check for updates

OPEN ACCESS

EDITED BY
Huajian Liu,
University of Adelaide, Australia

REVIEWED BY
Haoyu Lou,
University of Adelaide, Australia
David Legland,
Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

*CORRESPONDENCE
Yongjiang Zhang
yongjiangzh@sina.com
Xinyu Guo
guoxy73@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 09 May 2023
ACCEPTED 28 September 2023
PUBLISHED 13 October 2023

CITATION
Li Y, Huang G, Lu X, Gu S, Zhang Y, Li D, Guo M, Zhang Y and Guo X (2023) Research on the evolutionary history of the morphological structure of cotton seeds: a new perspective based on high-resolution micro-CT technology.
Front. Plant Sci. 14:1219476.
doi: 10.3389/fpls.2023.1219476

COPYRIGHT
© 2023 Li, Huang, Lu, Gu, Zhang, Li, Guo, Zhang and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Research on the evolutionary history of the morphological structure of cotton seeds: a new perspective based on high-resolution micro-CT technology

Yuankun Li[1,2†], Guanmin Huang[2†], Xianju Lu[2], Shenghao Gu[2], Ying Zhang[2], Dazhuang Li[2], Minkun Guo[2], Yongjiang Zhang[1*] and Xinyu Guo[2*]

[1]State Key Laboratory of North China Crop Improvement and Regulation, Key Laboratory of Crop Growth Regulation of Hebei Province, College of Agronomy, Hebei Agricultural University, Baoding, China, [2]Beijing Key Laboratory of Digital Plant, Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

Cotton (*Gossypium hirsutum* L.) seed morphological structure has a significant impact on the germination, growth and quality formation. However, the wide variation of cotton seed morphology makes it difficult to achieve quantitative analysis using traditional phenotype acquisition methods. In recent years, the application of micro-CT technology has made it possible to analyze the three-dimensional morphological structure of seeds, and has shown technical advantages in accurate identification of seed phenotypes. In this study, we reconstructed the seed morphological structure based on micro-CT technology, deep neural network Unet-3D model, and threshold segmentation methods, extracted 11 basics phenotypes traits, and constructed three new phenotype traits of seed coat specific surface area, seed coat thickness ratio and seed density ratio, using 102 cotton germplasm resources with clear year characteristics. Our results show that there is a significant positive correlation ($P<0.001$) between the cotton seed size and that of the seed kernel and seed coat volume, with correlation coefficients ranging from 0.51 to 0.92, while the cavity volume has a lower correlation with other phenotype indicators ($r<0.37$, $P<0.001$). Comparison of changes in Chinese self-bred varieties showed that seed volume, seed surface area, seed coat volume, cavity volume and seed coat thickness increased by 11.39%, 10.10%, 18.67%, 115.76% and 7.95%, respectively, while seed kernel volume, seed kernel surface area and seed fullness decreased by 7.01%, 0.72% and 16.25%. Combining with the results of cluster analysis, during the hundred-year cultivation history of cotton in China, it showed that the specific surface area of seed structure decreased by 1.27%, the relative thickness of seed coat increased by 8.70%, and the compactness of seed structure increased by 50.17%. Furthermore, the new indicators developed based on micro-CT technology can fully consider the three-dimensional morphological structure and cross-sectional characteristics among the indicators and reflect technical advantages. In this study, we constructed a

microscopic phenotype research system for cotton seeds, revealing the morphological changes of cotton seeds with the year in China and providing a theoretical basis for the quantitative analysis and evaluation of seed morphology.

KEYWORDS

cotton, seed morphological structure, micro-CT, phenotypic analysis, temporal succession

1 Introduction

Cotton is the main source of natural fiber and is widely cultivated both in China and worldwide due to its high economic importance (Ruan, 2005). The cotyledon, epicotyl, hypocotyl, radicle and seed coat are the structural components of cotton seed (Maeda et al., 2021). Additionally, cotton exhibits dark-colored pigment spots known as cotton phenol. The seed's morphological structure is closely associated with its functions, and differences exist in seed functions based on their various morphological structures. The size and shape of seeds are key agronomical traits that influence cotton yield and quality (Wu et al., 2022). Larger and fuller seeds demonstrate superior early growth, uniformity, and higher metrics such as single plant dry matter weight, root-to-shoot ratio, emergence rate, and leaf area (Liu et al., 1997; Wang et al., 2008). Additionally, the size and surface area of seeds can impact their water-holding capacity, rate of moisture absorption, and metabolic rate (Ozarslan, 2002; Feng et al., 2008). Further, there is a positive correlation between kernel-to-coat ratio, and seed oil content (Pahlavani and Abolhasani, 2006). A vibrant seed color, a prominent oil gland, and a full seed are indicative of greater vitality in cotton seed (Wang, 2007). The seed coat is the outermost protective layer of a seed, and its thickness is related to seed germination, drought, and other environmental stresses (Schüler et al., 2014). Knowing the relative thickness of the seed coat can help better comprehend the biological characteristics of seeds (Schüler et al., 2014). A seed cavity is a gas structure that is difficult to measure inside a seed. In a study of the subcutaneous cavity in maize, the subcutaneous cavity volume of maize seed was found to be one of the most significant factors affecting the grain breakage rate (Hou et al., 2019). However, there is no clear research on the formation and quantification of cotton seed cavities. Therefore, accurate measurement of seed morphological structures is of great significance for exploring seed functions and environmental adaptability. However, cotton seeds display irregular morphology. Traditional measurement methods, like using calipers to measure the linear dimensions of seeds or dissecting seeds to obtain internal physical parameters, produce unsatisfactory outcomes. Consequently, this impedes the progress of studying the morphological structure of cotton seeds. Therefore, agronomists urgently need accurate seed structure analysis methods to study the functional morphological relationships of seeds.

With the continuous advancement of agricultural digital technology and imaging technology, notable progress has been made in the analysis and evaluation of seed morphological structures. Regarding the methods used now for seed imaging, there are mainly four key aspects. The first part involves using two-dimensional (2D) images to examine seed shape. Capturing 2D images allows for the extraction of valuable information regarding the external morphology of seeds, including their size, shape, color and texture (Hacisalihoglu and Armstrong, 2023). 2D images commonly comprise a range of image types, such as red-green-blue (RGB) images, thermal imaging, and fluorescence images. Zhong et al. (2016) extracted the thickness index of wheat seeds using two-dimensional image light projection. In a study by Brodersen and Kuhl (2023) focusing on seaweed seeds, it was observed that photosynthesis in the seed sheath enhances the availability of oxygen in the central region of the seed under light exposure. This, in turn, increases respiratory energy production for biological synthesis and relieves internal oxygen deprivation within the seed. Additionally, Kranner et al. (2010) identified a significant correlation between temperature distribution changes and the seed imbibition process. Thermal imaging technology has demonstrated its viability as an alternative to conventional methods for assessing seed vigor. However, many agronomic traits need to be analyzed in a three-dimensional model (3D).

The second part focuses on the detection of the chemical composition of seeds using spectral technology. Spectral imaging is a fusion of spectroscopy and imaging technology. This method utilizes the absorption, scattering, or transmission characteristics of seeds to acquire chemical composition information within the seeds across varying wavelengths of light. This method offers several advantages, including high speed, high efficiency, non-contact, non-destructive, and reliable results. Commonly utilized hyperspectral images, multispectral images, near-infrared (NIR) imaging, and so on. In the current study, Fourier transform near-infrared (FT-NIR), dispersive diode array (DA-NIR) and hyperspectral imaging (HSI) have been successfully used to detect the quality of seed components such as protein, oil, water and starch. Furthermore, these methods have demonstrated the capability to predict chemical quality traits (Hacisalihoglu and Armstrong, 2023). Moreover, although fluorescence and thermal infrared images are not composed of a continuous spectrum of visible light wavelengths, they still involve the interaction between matter and light as well as the process of energy conversion of light. Therefore, they can also be classified as part of the spectrum. The combination of fluorescence and hyperspectral imaging also holds great potential for assessing heavy metal content (Zhou et al., 2023). Additionally, the fusion of proximal spectral phenotyping and 3D modeling registration has

emerged as a new development trend (Liu et al., 2020), but it has received limited attention in the field of seed research.

The third component is the 3D imaging technology using serial sections. Serial sections entail physically slicing the sample into thin slices. Next, each slice is observed and imaged under a microscope before layering the images to reconstruct the 3D structure of the entire sample. For instance, Ogawa et al. (2000) reconstructed the 3D morphological structure of rice seeds using consecutive sectional images. Commonly used 3D imaging techniques that make use of serial sections include traditional transmission electron microscopy (TEM) and scanning electron microscopy (SEM). This method provides detailed internal ultrastructural features of seeds, such as cell morphology and number (Rashid et al., 2020). In a study examining protein and vacuole formation in pea seed cotyledons, researchers reconstructed serial sections of tissues obtained on the 12th and 15th days, providing evidence for the possible formation of protein bodies (Craig et al., 1979). In a distinct study, Bhawana et al. (2014) utilized focused ion beam-scanning electron microscopy (FIB-SEM) to capture the 3D features of aleurone cells in Arabidopsis seeds. Consecutively slicing experimental samples is a time-consuming and destructive process that can lead to the exclusion of specific regions of interest (Prior et al., 1999). Moreover, incorrect thickness intervals between slices can cause deformation in the sections (Yamane et al., 2022).

The final part is 3D imaging technology. Commonly used 3D imaging techniques in seed research comprise magnetic resonance imaging (MRI), X-ray micro-computed tomography (Micro-CT, also known as μCT), confocal microscopy, and structured light imaging technology. 3D imaging technologies have the capability to examine anatomical internal structures without the need to slice them (Mohoric et al., 2009). Micro-CT utilizes X-rays for scanning and reconstructing the 3D structure of samples. This technique can provide high-resolution images, making it valuable for analyzing the internal structure of seeds (Yu et al., 2022). MRI employs magnetic fields and radio waves to generate high-contrast images containing anatomical information. MRI can offer insights into the internal structure and water distribution of seeds in seed research (Metzner et al., 2015). A 3D microscope employs lasers or near-infrared light sources for scanning samples and generating 3D models of seeds using stacked images. Structured light imaging involves using structured patterns generated by light sources to obtain the 3D external shape of samples. Despite having relatively larger errors compared to other techniques (Yu et al., 2022), structured light imaging still has certain applicability in studying the external morphology of seeds. In principle, MRI signals are based on the flow of water, while X-ray tomography is based on the differential absorption of X-rays by the sample, but compared with MRI, micro-CT can achieve higher spatial resolution (Metzner et al., 2015). Laser confocal microscopy (LSM) can nondestructively observe and analyze the internal microstructure of cells by constructing a 3D model of samples through 3D layer imaging, but it is slower than micro-CT imaging (Gubatz et al., 2007; Fanuel et al., 2018). So it seems appropriate to use micro-CT methods to explore seed structures.

In recent years, the application of micro-CT technology has become increasingly prevalent in the analysis of crop seeds. This innovative technology has been used to analyze the 3D morphological structure of seed organs in crops like rice, wheat, sorghum and maize. For example, micro-CT technology was used to measure the chalkiness index and quantify the crack size of rice seeds, promoting the genetic analysis of rice chalkiness regulation and quality evaluation in production (Su and Xiao, 2020). Le et al. (2019) assessed the morphology of wheat grain and its different compartments, quantifying the crease shape for each grain. Crozier et al. (2022) extracted phenotypic quantitative data from sorghum grain, including embryo volume, endosperm hardness, endosperm texture, endosperm volume, pericarp volume, and seed kernel volume, to enable the identification of genotypes with superior structural characteristics. Additionally, Yin et al. (2021) used micro-CT technology to analyze the variation of maize kernels from base to top, revealing the positional effect on grain growth and development. Overall, micro-CT technology offers a unique 3D perspective on characterizing seed morphological structure and vast potential for exploring the relationship between seed morphological structure, quality, and water transport (Warning et al., 2014). This technology represents a robust tool for studying the relationship between seed morphological structure and function and exploring the mechanisms of seed quality control. However, although this technology has been widely used on various crops, the micro-CT-based analysis technology of cotton seed has not yet been established, and the 3D structural phenotypic information on cotton seed remains undefined.

Over time, different breeding goals and specific production issues have resulted in varying growth characteristics and ecological adaptability of plant varieties, which can be reflected in significant differences in crop phenotypes. Seed phenotype plays a crucial role in contributing to crop phenotypes and reflects the profound genetic changes caused by intentionally or unintentionally human selection. For example, significant changes have occurred in the morphology and size of wheat grains during domestication and breeding due to the demand for flour protein particle content and hydrolytic enzyme activity (Gegas et al., 2010). This change has led to the classification of wheat cultivars, such as common wheat, hard wheat, cone wheat, and dense wheat. The shape of grains is divided into four categories: angular, oval, cylindrical, and elliptical, based on the shape of wheat seed kernels. Production-wise, wheat is often divided into two types based on the color of the grain: red skin and white skin. The red-skinned variety has a thicker skin, poor ventilation, and long dormancy, but is resistant to grain sprouting (Lang et al., 2021). However, seed morphological structure evolution is often ignored, and more focus is given to plant architecture evolution because of the similarity between germplasm. Cotton seeds are mainly identified based on features such as color and the chemical composition content of the seed kernel, as there is no clarity on the evolutionary characteristics of the seed morphological structure of cotton.

Given the absence of a 3D structure in cotton seeds and the challenge of accurately quantifying the internal structure, along with the unclear understanding of seed evolution, this study collected representative seed samples from different years and employed micro-CT technology to analyze the morphological structure of the

seeds. The results elucidated the process and trend of changes that occurred in the cotton seed throughout the different years. It was hypothesized that (1) the parameters of the micro-CT equipment can resolve the 3D morphological structure of the cotton seed; (2) there is a significant correlation around the morphological indicators of the cotton seed; and (3) the morphology of the cotton seed has distinct temporal characteristics as breeding progresses. The purpose of this study is to establish a micro-CT-based method for obtaining the morphology of the cotton seed, analyze the internal morphological characteristics of cotton, and investigate the evolution of cotton seed morphological structure. This study provides a high-throughput method for assessing the morphology of the seed morphological structure and a theoretical basis for the quantitative evaluation and analysis of the seed morphological structure.

2 Materials and methods

2.1 Experimental materials

The experiment was conducted at the Digital Plant Laboratory of the Beijing Academy of Agriculture and Forestry Sciences in 2022. The study incorporated 102 diverse types of cotton seeds (Table S1). 83 of these seeds were cotton varieties with explicitly identifiable breeding years and were grouped into four years based on the breeding timeline (Yu, 2018). Years 1-4 pertain to cotton varieties that were introduced from overseas or domestically cultivated and widely promoted in China within the years 1904-1958, 1958-1970, 1970-1990, and 1990-2020, respectively.

Among them, 51 seeds and 5 seeds were sown and harvested in Hebei Province and Xinjiang Province, respectively, in 2020. Sixty-one samples were stored in the national cotton germplasm resources Mid-Term Genebank at a temperature of 0°C ± 2°C and a relative humidity of 50% ± 7%.

2.2 Cotton seeds retrieving and analyzing methods establishing

2.2.1 Micro-CT image acquisition

The micro-CT device used for this study was the SkyScan 1172 (Bruker, USA). The scanning resolution was set at 2K (2000×1332), with a scan pixel pitch of 12.86 μm and an angular step of 0.400°. The scan was conducted using a voltage of 40 kV and a current of 250 μA, and each individual scan required approximately 29-30 minutes to complete. To improve the accuracy and efficiency of scanning, three cotton seeds were secured on a homogeneous foam panel, and the panel was fixed to the rotary table of the micro-CT scanning equipment to decrease the influence of other fixed matrices on image grayscale. After each scan, 488 images (. TIF) were acquired for each variety, resulting in a total size of 2.41 GB. The CT Scan NRecon software (Bruker, US) was utilized for image reconstruction. Each reconstructed image had a format of BMP with a resolution of 2000×2000px, and their sizes ranged from approximately 3 to 4 GB.

2.2.2 Single seed segmentation

First, the original image was downsampled by a factor of 3. The threshold for image binarization was automatically determined using the Otsu algorithm (Otsu, 1979). Morphological operations were employed to eliminate internal holes and generate an image mask. Next, the three-dimensional watershed algorithm (Neubert and Protzei, 2014) was utilized to segment individual seed kernels. To extract micro-CT images of individual seed kernels, the segmented image was combined with the original image using the "AND" operation.

2.2.3 Seed kernel phenotyping pipeline

In the micro-CT images of cotton seeds, there is high pixel grayscale in their seed kernels and seed coats. To enhance the efficiency of seed image processing, we aim to develop a phenotyping pipeline. The dataset we used comprises 286 scanned images of 20 seed kernels. Next, we employed the effective interactive segmentation (EISeg) method described by Hao et al. (2022) to manually label the embryonic region of these images (Figure 1A). Among these structures, the seed kernel is more easily identifiable and labelable. However, manually labeling the seed coat and cavity structure resulted in significant errors. Therefore, we only labelled the kernel for model training. Additionally, the data was divided into training and testing sets with a ratio of 8:2, chosen randomly.

We utilized the PyTorch - cpu0.8.2 deep learning framework, along with libraries such as SimpleITK, numpy, scipy, skimage, and vedo in Python (3.9.0). The U-net 3D network architecture was illustrated in Figure 1B. This network consisted of an encoder and a decoder, taking in images of size 32× 128× 128 and producing seed kernel masks of size 32× 128× 128. The network was divided into 4 layers, performing 5 down sampling operations during the encoding phase and 5 up sampling operations during the decoding phase. Each layer was composed of two Conv3D-BN-ReLU modules connected by residual connections. The number of convolutional kernels in each layer was 16, 32, 48, 64, and 96, respectively. MaxPool and ConvTranspose were used for downsampling and upsampling, respectively. LeakyReLU was employed as the activation function in the activation function and normalization layer, while InstanceNorm was used as the normalization layer to enhance the model's expressive power. To restore the low-resolution feature maps to the original image resolution, bilinear interpolation was applied in the upsampling layer. Furthermore, the model's output was normalized and probabilized using the Softmax and Sigmoid functions. The model configuration included setting the Epoch to 300, the learning rate to 0.0001, the Batch_size to 4, and performing the training using the Adaptive Momentum Estimation (Adam) optimization algorithm. The loss function used was Dice_Loss. The accuracy of the network was evaluated using the Dice similarity coefficient (DICE) and Intersection over Union (IoU).

During both the training and inference stages of the model, the input images were processed in a chunking manner. In the training stage, 32× 128× 128 volumetric data samples were randomly extracted from the original images and their corresponding labels as inputs to the network (Figure 1B). During the inference stage,

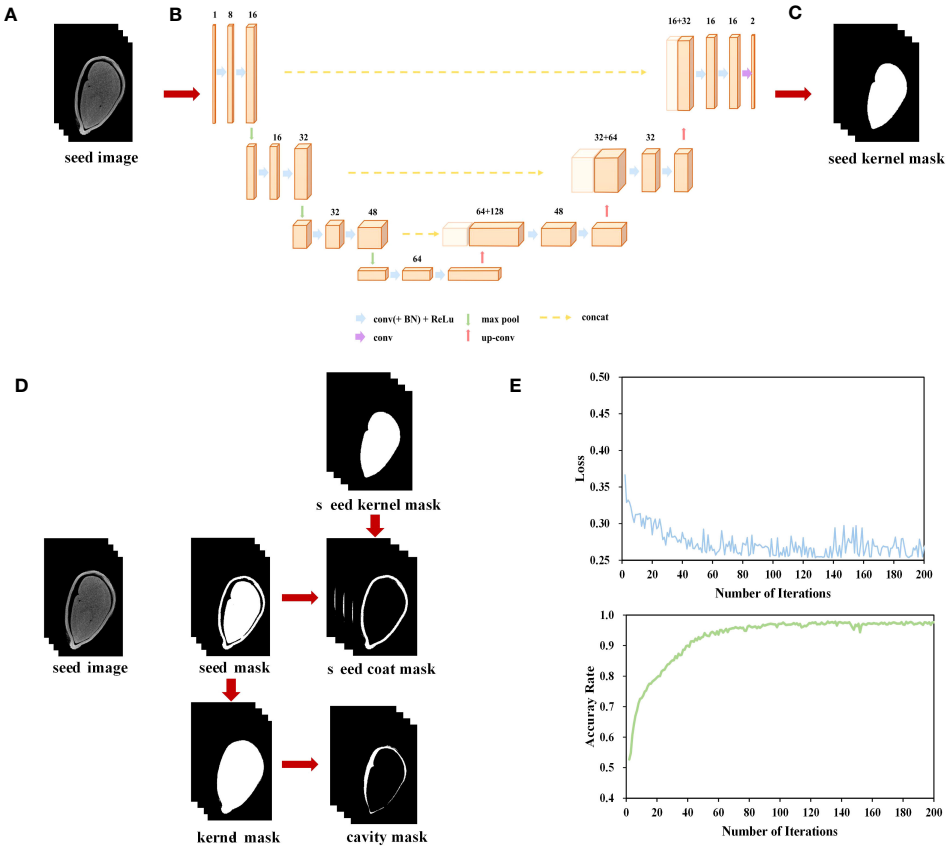


FIGURE 1
Micro-CT-based cotton seed imaging processing flow. **(A)** micro-CT image sequence. **(B)** U-net network. **(C)** Segment seed kernel mask. **(D)** Segment seed coat and cavity. **(E)** Loss and accuracy curves on training set.

volumetric data of size 32 ×128 ×128 was sequentially extracted with a stride of (24, 96, 96), and then fed into the segmentation network for inference, ultimately obtaining the seed kernel mask image (Figure 1C).

Throughout 200 training iterations with a learning rate of 0.0001, the loss consistently decreased, converging to a model's accuracy of 97.7% (Figure 1E).

2.2.4 Seed coat and cavity extraction

We utilized the Otsu algorithm (Otsu, 1979), a basic image processing technique, to segment the seed coat and cavity based on the disparity in gray levels between the target object and the background in the given image. We performed a gray-level histogram analysis on each individual seed kernel to determine the threshold for binarization and obtain the seed mask image. Subsequently, this image was utilized to create a mask encompassing the contour of the seed. Finally, by utilizing the "AND" operation, the masks corresponding to the seed coat and cavity were merged with the original image, resulting in the production of seed coat and cavity images (Figure 1D).

Open-source medical imaging processing tools, including Itk-SNAP (Yushkevich et al., 2006) and 3D Slicer (www.slicer.org) (Fedorov et al., 2012), were employed for rendering and 3D visualization purposes.

2.3 Sampling and measurements

2.3.1 Micro-CT analysis

Based on the target mask image and the target surface model reconstructed by the Marching Cube algorithm (Lorensen and Cline, 1987). Based on the surface model, 11 phenotypic traits of cotton seeds, comprising seed, seed coat, seed kernel, and cavity morphological structure (Table S2) were extracted. It is important to mention that the seed surface area is also known as the seed coat surface area, and the usage of these terms in the paper depends on the context of the paragraph.

2.3.2 Manual measurement

In order to compare the measured dimensions of seeds with their corresponding extracted data, it is necessary to use a vernier caliper to measure and record their length, width, and thickness. We defined length as the maximum dimension of the seed, while width is the maximum dimension perpendicular to its length (Hu et al., 2018).

We defined the thickness as the straight-line size perpendicular to both the length and width directions (Hu et al., 2018).

2.4 Seed morphological structure evaluation indicators

The seed coat specific surface area is expressed by the ratio of the seed coat surface area to its volume, which can reflect the surface area of the seed coat per unit volume. The larger the specific surface area of the seed, the greater its exposure to the surrounding environment.

$$\text{Seed coat specific surface area } (m^2/m^3) = \text{Seed coat surface area}(m^2)/\text{Seed coat volume}(m^3) \quad (1)$$

The seed coat thickness ratio is defined as the average seed coat thickness divided by the seed thickness. The seed coat thickness ratio represents the relationship between the thickness of the seed coat and that of the seed, with a higher value indicating thicker seed coat.

$$\text{Seed coat thickness ratio} = \text{Average seed coat thickness (m)}/\text{Seed thickness (m)} \quad (2)$$

The seed density ratio is defined as the ratio of the cavity volume of to the seed kernel volume. It is commonly used to evaluate the internal morphological structure of seeds, with a smaller SDR value indicating a denser internal structure of the seeds.

$$\text{Seed density ratio} = \text{Cavity volume}(m^3)/\text{Kernel volume}(m^3) \quad (3)$$

2.5 Data analysis

Experimental data were organized using Microsoft Excel 365 (Microsoft Corporation, USA) with statistical analyses carried out using SPSS Statistics 25 (IBM Corporation, USA) for variance analysis and variety clustering. ANOVA analysis of variance (generalized linear model) was used. When the data conformed to a normal distribution, we used LSD multiple comparisons for data that conforms to normality. For data that did not conform to normality, in order to avoid false positive results, we used Bonferroni multiple comparison results. The least significant difference (LSD) method was applied for multiple comparisons, with significant differences between different seed parameters compared on a $P<0.05$ level. The average seed coat thickness indicator was excluded from the clustering analysis due to its insignificant change. Therefore, ten seed phenotypic indicators were standardized using the Z-score method and classified using the Ward method in combination with squared Euclidean distance as the similarity measure, to categorize the indicators of different cotton varieties.

3 Results

3.1 Establishment of micro-CT acquisition and analysis method for cotton seeds

Since cotton is a dicotyledonous crop, the cotyledons of cotton seeds are curled and closely connected to the radicle and hypocotyls. Meanwhile, distinguishing between the radicle, hypocotyl, primordial epicotyl, and two cotyledons was difficult in our micro-CT images. Thus, these anatomical components were regarded as a unified morphological structure known as the seed kernel for the purpose of segmentation and computational analysis. Figure 2 presented 3D reconstruction images of the 102 cotton varieties and the seed coat, cavity, and seed kernel of 12 cotton varieties from various angles.

Upon observing the micro-CT scan images of cotton seeds, it became evident that the cavity structure occurs between the seed kernel and seed coat (Figure 3A). Three other types of cavity morphological structures also existed, namely the internal cavity of the seed kernel (Figure 3B), the cavity between the endosperm residue and the seed coat (Figure 3C), and the cavity between the internal and external seed coat. However, unlike other crop seeds, the latter three cavity morphological structures of cotton seeds are atypical and mainly occur in dry, dehydrated, or even dead seeds. Thus, this article solely focuses on the cavity between the seed kernel and seed coat.

3.2 Analysis of phenotypic indicators of cotton seeds

Eleven phenotypic traits of cotton seeds were obtained through the analysis of micro-CT images. Descriptive statistics showed that the average seed length of 102 cotton seeds was 9.22 mm, with a relatively small standard deviation and coefficient of variation values (Table 1). The average values for seed width and thickness were also stable. The thinnest and thickest seed coats had average thicknesses of 0.10 mm and 0.21 mm, respectively, but relatively larger coefficient of variation values than seed length, width, and thickness. Kernel volume and kernel surface area had larger coefficient of variation values, with values of 0.21 and 0.14, respectively (Table 1). On average, the seed kernel accounted for 54% of the total seed volume, with the remaining 46% comprising the seed coat and internal cavity. Notably, the coefficient of variation of the cavity was as high as 0.71 (Table 1). Comparing the data obtained through micro-CT with those obtained through manual measurements, the coefficients of determination for seed length, width, and thickness were 0.87, 0.83, and 0.81, respectively (Figure 4).

Through Pearson correlation analysis of 11 phenotypic characteristics of cotton seeds (Figure 5), it revealed a highly significant positive correlation between the volume and surface area of the seed and the volume and surface area of the seed kernel, as well as the volume of the seed coat ($r = 0.57$~0.83, $P< 0.001$). There was a weak correlation between the cavity volume and the seed coat volume ($r = 0.37$, $P< 0.001$), and a weak correlation with other phenotypic characteristics ($r < 0.30$, $P < 0.001$) (Figure 5). Additionally, the volume and surface area of the seed kernel showed a positive correlation with the volume and surface area of the seed and with the volume of the seed coat ($r = 0.51$~0.83, $P< 0.001$) (Figure 5). In addition, seed coat volume was positively correlated with the volume and surface area indexes of other phenotypic features ($r = 0.51$~0.74, $P<0.001$) (Figure 5). However, the seed thickness was moderately and

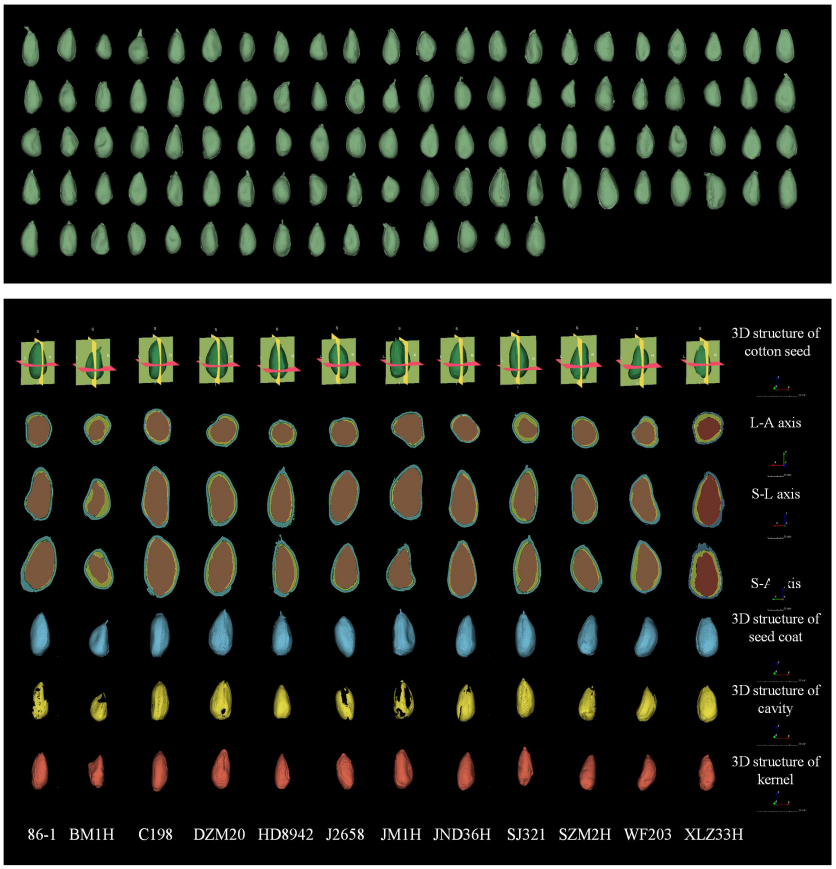


FIGURE 2
Three-dimensional reconstruction of 102 cotton seed (green) and three-dimensional reconstruction images and three-dimensional view of seed coat (blue), cavity (yellow), and seed kernel (red) of 12 seeds. The seeds shown in the figure are represented using the RAS coordinate system and demonstrate consistent orientation, measured in mm. The three-dimensional morphological structure was scaled to 25 cm, and the image scale for the three orthographic views was set to 5 cm.

positively associated with the seed coat volume ($r = 0.60$, $P < 0.001$), and had a weak correlation with the seed volume and cavity ($r< 0.03$, $P< 0.001$), and was not significantly correlated with other phenotypic characteristics (Figure 5).

3.3 Changes trends in cotton seed morphological structure in different years

Comparisons of seed phenotypic characteristics from different years (Figure S1), from 1904-1958 to 1958-1970, showed significant changes in length, width, seed volume, seed kernel volume, seed surface area, seed coat volume, cavity volume, and seed fullness. Conversely, the average seed coat thickness had a minimal and insignificant decrease (Figure S1). 1958-1970 had a significantly smaller length (8.56 mm), thickness (4.71 mm), seed volume (88.12 mm^3), seed kernel volume (55.28 mm^3), and seed coat volume (20.19 mm^3) than the other years (Figure S1). From 1958-1970 to 1970-1990, all indicators, except for seed fullness, showed an upward trend, and most indicators had significant differences compared to 1958-1970 (Figure S1). However, when comparing the indicators of 1904-1958, 1958-1970, and 1970-1990, 1904-1958 and 1970-1990 had less significant differences in the various indicators (Figure S1). In 1990-2020, the degree of decline in length, width, thickness, seed surface area, seed coat volume, average seed coat thickness, and seed fullness compared to 1970-1990 was not very significant (Figure S1). Nevertheless, there was a remarkable decrease in seed volume, seed kernel volume, and seed kernel surface area, by 8.56%, 14.11%, and 7.82%, respectively (Figure S1).

Comparing the changes in the seed morphological structure of varieties domestically bred in China from 1958 to 2020 (Figure 6), we observed an upward trend in seed volume and surface area, seed coat volume, cavity volume and average seed coat thickness of domestically bred varieties. They increased by 11.39%, 10.10%, 18.65%, 115.76%, and 7.85%, respectively (Figure 6). However, the seed kernel volume, seed kernel surface area, and seed fullness showed a downward trend, decreasing by 7.01%, 0.72%, and 16.25%, respectively (Figure 6). Among them, the seed coat volume and average seed coat thickness had goodness-of-fit values of 0.70 and 0.88, respectively (Figure 6). The cavity volume had an R^2 value of 0.9671, and the goodness-of-fit of seed fullness was the highest, approaching 1, indicating high predictability (Figure 6).

10.3389/fpls.2023.1219476

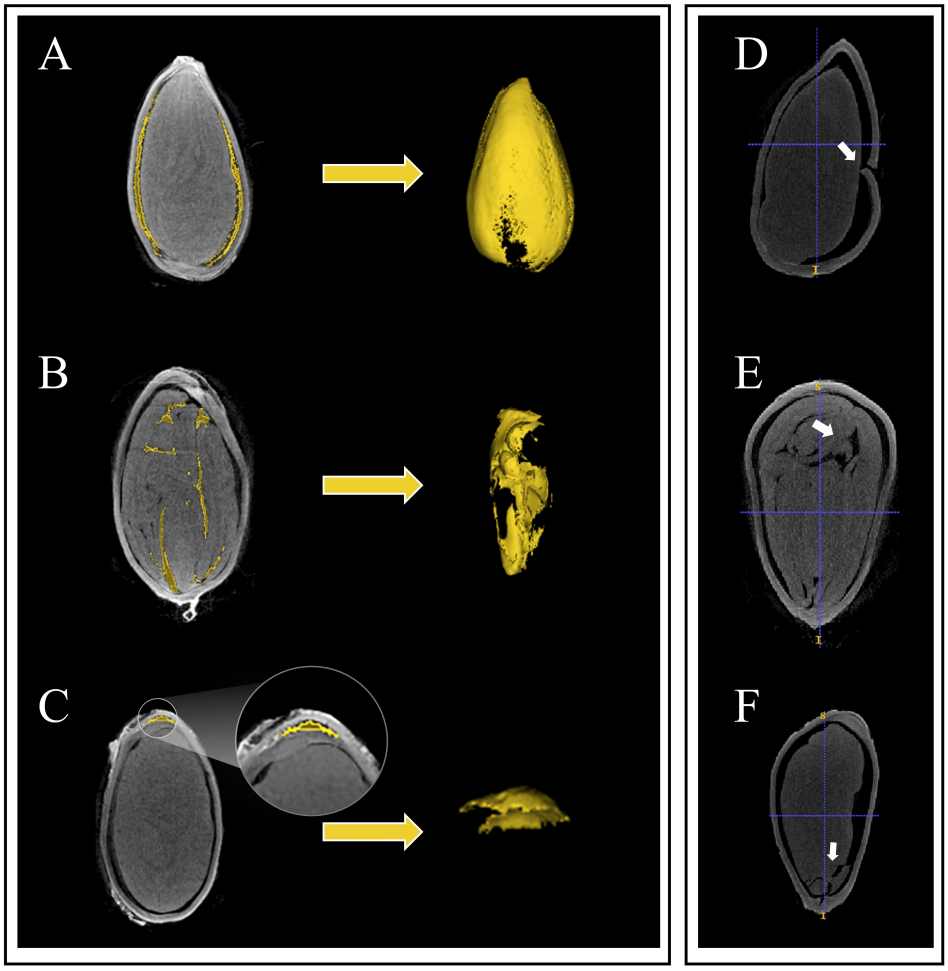


FIGURE 3
Three types of the inner structures and seed damage in cotton seeds are depicted. Among them, **(A)** represents the cavity between the seed coat and the kernel, **(B)** depicts the cavity inside the seed coat and endosperm remnants, **(C)** shows the cavity inside the seed kernel, **(D)** represents damage to the seed coat, **(E)** shows the breach between seed kernels, and **(F)** represents damage to the seed kernel.

TABLE 1 Descriptive statistics for the phenotypic indicators of 102 cotton seeds.

Morphological trait	Unit	Mean ± SD [a]	CV [b]	Range
Seed Length	mm	9.22 ± 0.70	0.08	7.26-11.16
Seed Width	mm	5.40 ± 0.50	0.09	4.44-8.49
Seed Thickness	mm	4.86 ± 0.41	0.08	3.67-6.91
Seed Volume	mm^3	100.34 ± 15.40	0.15	63.3-150.71
Seed Surface Area	mm^2	148.51 ± 21.14	0.14	106.52-223.83
Seed Kernel Volume	mm^3	54.25 ± 11.59	0.21	26.82-95.37
Kernel Surface Area	mm^2	89.26 ± 12.51	0.14	61.08-129.92
Seed Cavity Volume	mm^3	10.33 ± 7.31	0.71	0.09-41.94
Seed Coat Volume	mm^3	23.99 ± 4.27	0.18	12.96-41.27
Average Seed Coat Thickness	mm	0.16 ± 0.02	0.12	0.1-0.21
Seed Fullness	%	0.54 ± 0.06	0.12	0.3-0.67

[a] SD is the abbreviation for standard deviation; [b] CV is the abbreviation for Coefficient of variation.

10.3389/fpls.2023.1219476

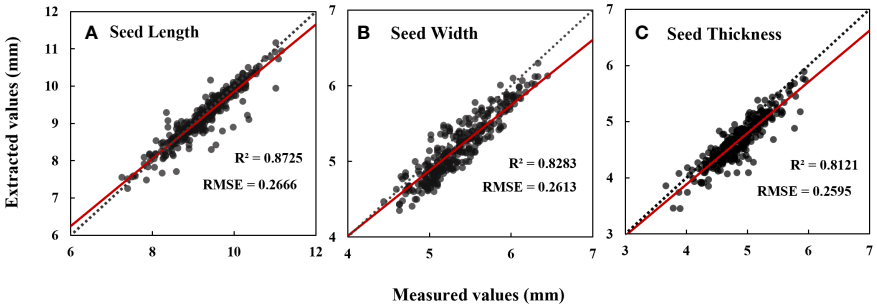


FIGURE 4
Data evaluation of seed length **(A)**, seed width **(B)** and seed thickness **(C)** measured values extracted based on CT images. N=306. Date represents mean ± SE (3 biological replicates, n=9, 15 and 26 plants, respectively), letters above the bars indicate significant differences at the level of P<0.05.

3.4 Similarity and classification of cotton seeds from different years

A variance analysis was performed on all seed phenotypic indicators, which demonstrated that these indicators had statistical significance in all examined varieties. Nonetheless, the seed coat thickness indicator was excluded from the clustering analysis due to its insignificant change. Ten seed phenotypic indicators were standardized using the Z-score method. After standardizing all phenotypic data, we performed Ward clustering, resulting in the classification of 102 varieties into three major clusters based on the squared Euclidean distance (Figure 7).

LSD test was conducted on three groups of 102 varieties (Figure 8), and the results showed that group-1 was classified as a small seed group with the smallest seed volume and seed kernel volume and the largest cavity volume. The seed size of the group-2 was at a medium level, the cavity volume was slightly lower than that of the group-1, and the seed coat volume was significantly higher than that of the group-3, so it could be classified as the middle seed group. Group-3 had the largest seed volume and seed kernel volume, but the smallest cavity volume, which is classified as the large seed group.

The 1904-1958 group exhibited an even distribution of varieties among large, medium, and small seed clusters, at 20%, 40%, and 40%, respectively (Figure S2). In the 1958-1970 group, 75% of the varieties was distributed in small seed clusters, while the remaining 25% were distributed in large seed clusters (Figure S2). The 1970-1990 group showed a roughly similar distribution among three seed clusters, accounting for 36.4%, 27.2%, and 36.4% (Figure S2). For

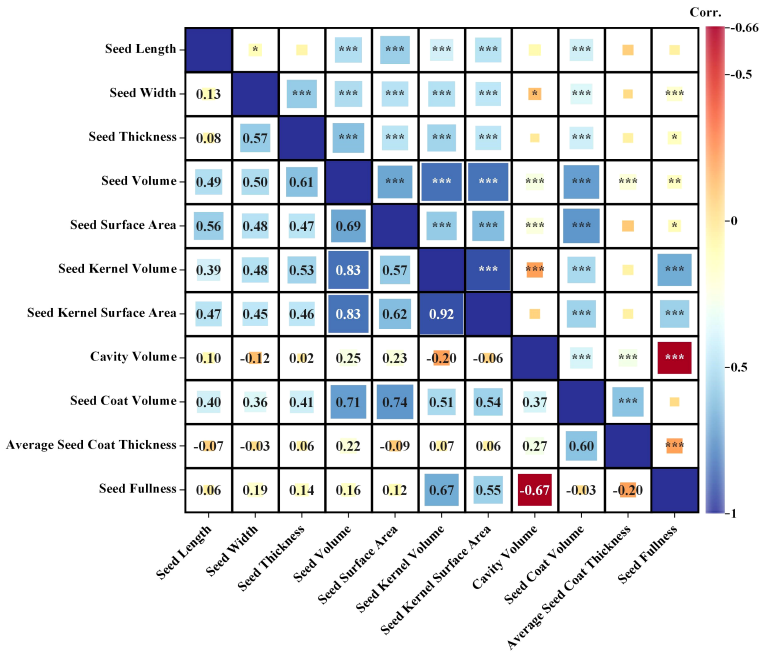


FIGURE 5
Correlation analysis of 11 phenotypic indicators (Seed Length, Seed Width, Seed Thickness, Seed Volume, Seed Surface Area, Kernel Volume, Kernel Surface Area, Seed Coat Volume, Seed Cavity Volume, Average Seed Coat Thickness and Seed Fullness). Significance *< 0.05, **< 0.01; ***< 0.001.

10.3389/fpls.2023.1219476

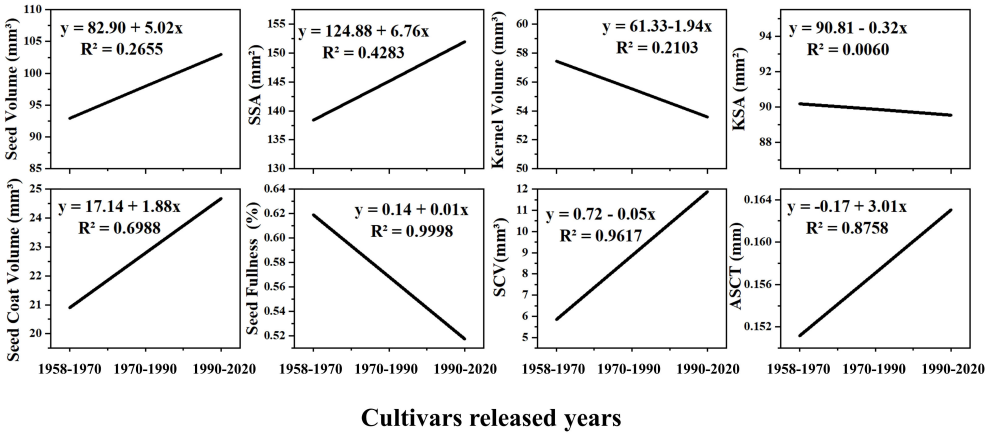


FIGURE 6
The trend of seed morphological structure change of cotton varieties independently cultivated in China. SSA, Seed Surface Area; KSA, Seed Kernel Surface Area; SCV, Seed Cavity Volume; ASCT, Average Seed Coat Thickness. The cultivars released years involve cotton varieties cultivated from the country in 1958-1970, 1970-1990 and 1990-2020, respectively.

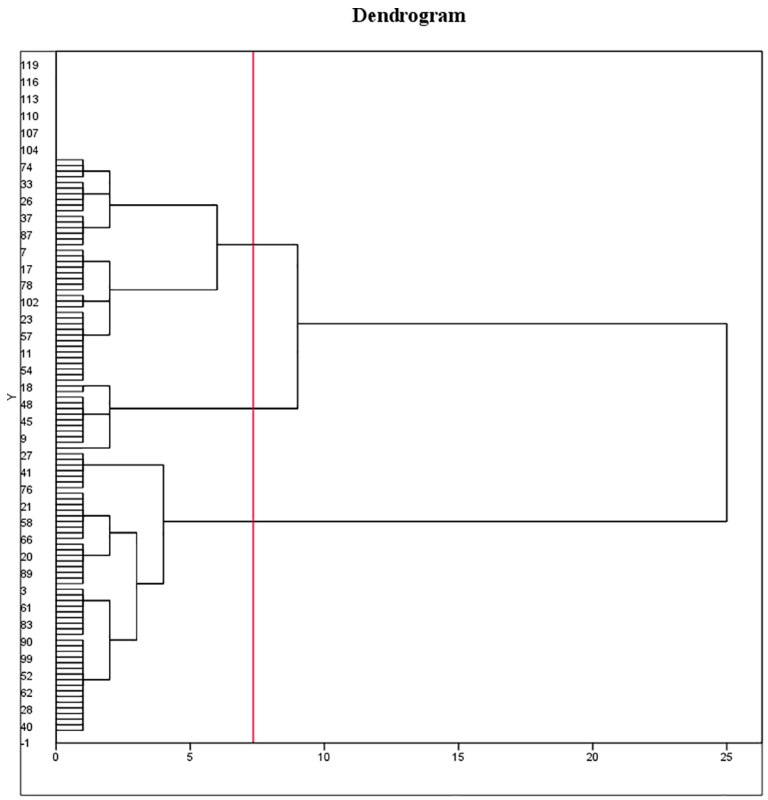


FIGURE 7
Clustering of 102 cotton varieties. Ten seed phenotypic indicators were standardized using the Z-score method and classified using the Ward method in combination with squared Euclidean distance as the similarity measure, to categorize the indicators of different cotton varieties. The 102 varieties are divided into three categories. The first cluster comprised 46 varieties, accounting for 45.1% of the total varieties. Of these, one variety was from 1904-1958, three varieties from 1958-1970, four varieties from 1970-1990, 32 varieties from 1990-2020, and six varieties had unknown Years. The second cluster consisted of 41 varieties, accounting for 40.2% of the total varieties, including 2 varieties from 1904-1958, 3 varieties from 1970-1990, 30 varieties from 1990-2020, and 6 varieties with unknown Years. The third cluster encompassed 15 varieties, accounting for 14.7% of the total varieties, including 2 varieties from 1904-1958, 1 variety from 1958-1970, 4 varieties from 1970-1990, 3 varieties from, and 5 varieties with unknown Years.

10.3389/fpls.2023.1219476

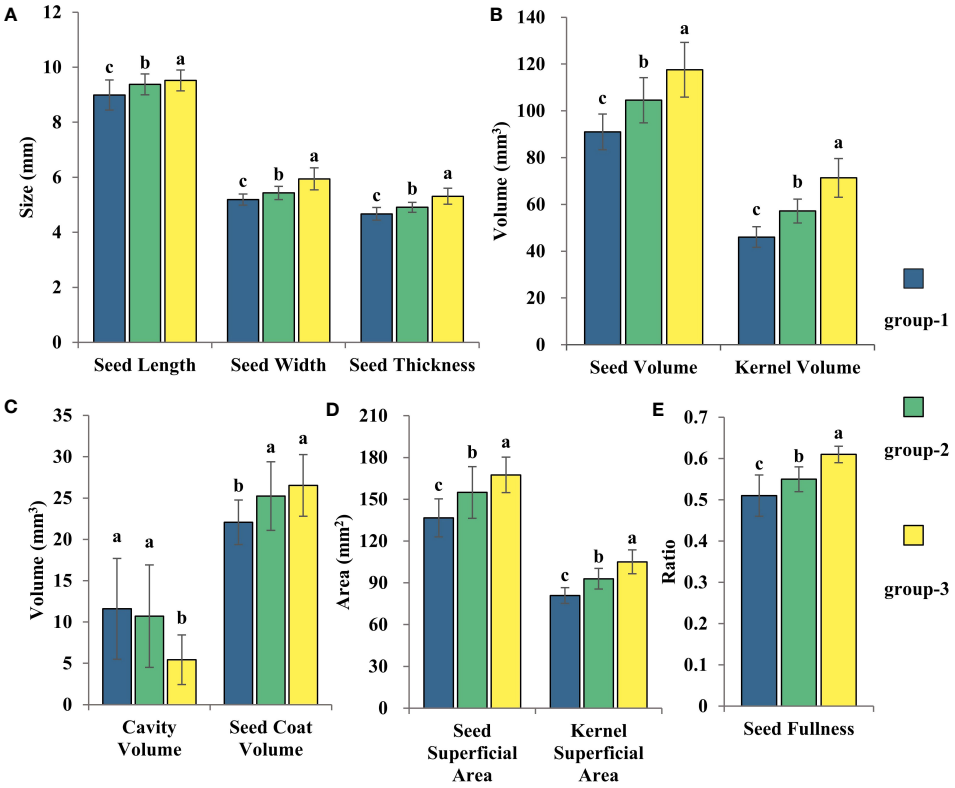


FIGURE 8
The three clustering results from the statistical analysis of the variations in cotton seed morphology were Seed Length, Width, and Thickness **(A)**, Seed Volume and Seed Kernel Volume **(B)**, Cavity Volume and Seed Coat Volume **(C)**, Seed Surface Area and Seed Kernel Surface Area **(D)**, and Seed Fullness **(E)**. LSD test was used for normal distribution data. Date represents mean ± SE (3 biological replicates, n=46,41 and 15 varieties, respectively), letters above the bars indicate significant differences at the level of P<0.05.

the 1990-2020 group, the majority of the varieties were distributed in small and medium seed clusters at 50.8% and 44.4%, respectively (Figure S2). Only 4.8% of the varieties was distributed in the large seed cluster (Figure S2). Thus, between 1958-2020, the seed morphology of self-bred varieties in China transitioned from first changing from large and small seed groups to three seed groups of large, medium and small, and finally to the change process of middle and small seed groups.

3.5 Evaluation of seed morphological structure in different years

In order to facilitate better evaluation of morphological structural differences among different varieties throughout time, this study proposed three specific parameters, namely seed coat specific surface area, seed coat thickness ratio and seed density ratio. The evaluation of seed coat specific surface area among distinct generations of cotton seeds, indicated in Figure 9A, illustrates that respective average roughness measurements were 6.31 m^{-1}, 6.66 m^{-1}, 6.23 m^{-1}, and 6.28 m^{-1} for the 1904-1958 through 1990-2020. While the seed coat specific surface area increased initially from 1904-1958 to 1990-2020, it underwent a decreasing trend afterwards, but there was no significant variation in seed coat specific surface area among these generational cohorts. Moreover, implications derived from seed coat thickness ratio analysis (Figure 9B) showed a low degree of characterization of seed coat thickness ratio among these distinct years, as seen by the mean value ranging between 0.031 and 0.034 showing an increasing trend — that was, the relative thickness of the seed coat increased. Calculation from Figure 9C demonstrated that the mean value of the seed density ratio elevated from 0.15 to 0.23 throughout the 70-year time period. However, for 1958-1970, the seed density ratio was set at 0.10, which is considerably below the recorded values of the remaining years. On the whole, the results of the seed density ratio showed that the seed morphological structure in self-bred varieties in China was becoming more compact. While the cavity was increasing year by year, the seed kernel size was increasing at a faster rate than the cavity. The seed thickness ratio and specific surface area of the seed coat exhibited an upward trend. However, the comparative approach, taking into account the overall morphological structure of the seed, offered a more compelling depiction of the average change in seed coat thickness.

The correlation analysis of the three parameters, presented in Table 2, identified a significant association among them. Of these parameters, seed coat specific surface area and seed coat thickness ratio showed the highest correlation coefficient (r = -0.80, P< 0.001), suggesting the likelihood of their representing similar traits. Hence, when assessing seed morphological structure, either of the parameters could be utilized without necessitating the requirement to utilize both.

10.3389/fpls.2023.1219476

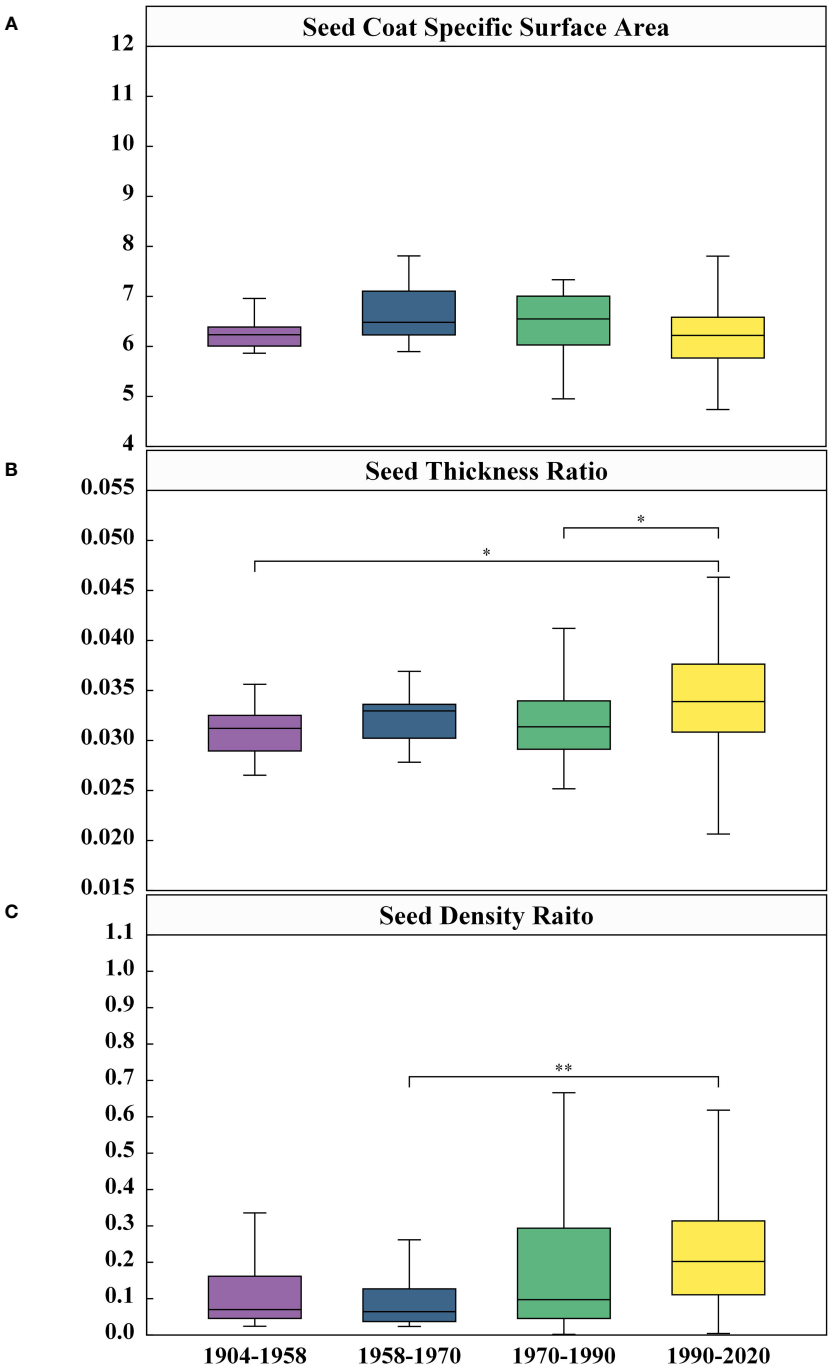


FIGURE 9
Comparison of differences in Seed Coat Specific Surface Area **(A)**, Seed Coat Thickness Ratio **(B)**, and Seed Density Raito **(C)** among four Years. Given are the means ± SEM. Boxes represent first and third quartile (upper and lower margins), and median (horizontal line). N=306, significance: *<0.05, **<0.01.

4 Discussion

4.1 Analyzing the microscopic phenotype characteristics of seeds

Micro-CT nondestructive imaging technology has been widely used in plant phenomics (Guelpa et al., 2015; Hu et al., 2020; Wu et al., 2021). This paper presented a novel technique for studying the microscopic phenotype of cotton seeds. In most studies, researchers obtained a large number of micro-CT images from single seeds by scanning micro-CT and processing them individually by manual frame selection or commercial software, which took a lot of labor time (Hou et al., 2019; Dong et al., 2020). Therefore, this paper aims to transform the original batch seed kernel micro-CT image

TABLE 2 Correlation analysis of four seed evaluation indicators.

	Seed Coat Specific Surface Area	Seed Coat Thickness Ratio	Seed Density Ratio
Seed Coat Specific Surface Area	1	-0.80***	-0.26**
Seed Coat Thickness Ratio		1	0.26**
Seed Density Ratio			1

**Represents significant correlation at P<0.01, ***Represents significant correlation at P<0.001.

processing problem into a single seed kernel segmentation problem. Additionally, the analysis of individual seed kernel components is also transformed into a semantic segmentation problem, focusing on the seed kernel and cavity components that possess the most prominent internal characteristics. The method improved the efficiency of seed treatment and quickly extracted 11 seed structural traits. The micro-CT equipment took 29-30 minutes to scan a single seed, and scans three seeds each time. In contrast to the conventional commercial software utilized for single seed processing, the seed processing technique described in this paper is characterized by its brevity and enhanced efficiency. Consequently, a complete microscopic phenotype technique system for cotton seeds has been established, which provides significant technical support for furthering the exploration of the internal morphological structure of the seeds.

The contrast of the image obtained by micro-CT theoretically depends on the density, thickness and molecular structure of the sample (Lusic and Grinstaf, 2013). Owing to the communication of the seed cavity under a damaged seed coat with the external air (Figure 3D), obtaining accurate segmentation and quantification would be impracticable. This makes it difficult to distinguish grayscale pixels. Hence, in this study, only seeds with undamaged seed coats were utilized for accurate segmentation and quantification. Moreover, dehydration in seeds may result in the potential connection of the internal cavity of the seed kernel (Figure 3E) with the exterior cavity of the seed kernel, consequently rendering the task of obtaining accurate segmentation and quantification challenging. Furthermore, we also observed the situation of seed kernel breakage (Figure 3F), which is comparably infrequent. Given the unclear structure of the kernel, we segmented cotton seeds into three components: seed coat, cavity, and kernel. Here is a study that is similar to our segmentation results. In walnut, researchers only segmented the cavity, seed kernel and shell (Bernard et al., 2020). Conversely, in seeds of monocotyledonous crops, Hou et al. (2019) defined three types of cavities in maize seeds: the embryonic cavity, the endosperm cavity, and the subcutaneous cavity outside the endosperm. However, the seed coat and endosperm of maize seeds are fused together, so the authors treated them as a whole during calculations. Although this paper extracted two phenotypic indicators, seed coat volume and average seed coat thickness, inner and outer seed coat features were not successfully segmented. Further research and exploration are needed to accurately segment seed morphological structures.

Wu et al. (2022) analyzed the correlations of seed traits including weight of 100 seeds, seed length, seed width, seed length-width ratio, seed area, seed perimeter, seed diameter, and sphericity with an automatic seed testing machine and found that these morphological traits are environmentally stable. In the correlation analysis conducted by Wu et al. (2022), the correlation between seed length and seed width was low, while the correlation between the seed width and seed thickness was high, which was consistent with our research results. Nevertheless, there was a strong positive correlation between seed length and thickness, as well as between seed surface area, width, and thickness, which contradicted the findings presented in this paper.

Seed kernel is the best source of seed protein and oil, which relates to seed size (Huang et al., 2022). It means that a larger seed kernel size may produce more nutrients. Our study found a positive correlation between seed kernel size and seed volume, which is consistent with previous research results (Bernard et al., 2020). In addition, the condition of the kernel is an important indicator of seed quality, and larger cotton seed kernels contribute to seedling growth (Ahmed et al., 2020). Clearly, seed kernel size is mainly influenced by seed coat and size; that is, the larger the seed, the larger the seed kernel. Micro-CT scanning of the seed's internal shape and structural characteristics is a reliable method to predict cotton growth.

In this study, the variation coefficient of cavity volume was too high. Correlation analysis of 102 cultivars indicated a positive correlation between seed coat volume and cavity volume, suggesting that the formation of cotton seed cavities may be related to the morphological structure of the seed coat. However, it is unclear whether this structural feature varies among cultivars, and further exploration is necessary to determine the reasons for the high variability coefficient.

Generally, the variability coefficient of average seed coat thickness was small across cultivars, and average seed coat thickness had no significant correlation with most phenotypes except for those related to seed coat volume. This is consistent with the findings of Bernard et al. (2020). Therefore, there may be unidentified factors that affect the average seed coat thickness.

4.2 Differences and evaluation of seeds in different years

Understanding the changing trend of seed size is critical for germplasm enhancement. Researching the evolution law of germplasm contributes significantly to the growth and regulation of cotton seeds and organs, and has vital implications in setting breeding objectives, selecting parent materials, and offspring (Huang et al., 2022).

The application of computed tomography scanning technology has revealed the correlation between small cavities in soybeans and their oil content. More small cavities are present in modern soybeans compared to ancient soybeans, suggesting a gradual upsurge in soybean oil demand (Zong et al., 2017). It indicates that micro-CT has significant potential to explore the crop domestication process. Our study classified 83 cotton varieties into four distinct periods and uncovered various alterations in seed morphological structure throughout the course of cotton development in China. Our research exhibited that there were notable variations in seed coats, seed kernels, and cavities amongst cotton seeds from different periods. However, the average seed coat thickness did not demonstrate a noticeable trend. There was a growing trend in the seed coat, kernel, and cavity sizes of self-bred varieties in China. This suggested that the focus of breeding bred varieties in China is on increasing seed and kernel sizes, with less emphasis on the average seed coat thickness.

This paper presents three innovative indicators to assess the morphological structure of the seeds. Our objective is to improve the understanding of the relationship between seed size, seed coat thickness, and internal morphological structure throughout the breeding process. Among them, the seed coat specific surface area of domestic self-bred varieties in China were showing a decreasing trend. This may be because with a larger specific surface area, water is absorbed more quickly for seed, and the rapid absorption of water can damage seed cells and affect the cleanliness of seedlings (Main et al., 2014). These indicators could help develop better breeding plans and strategies. The indicators proposed in this paper also reflected the differences in seed relative thickness and internal morphological structure compactness between China's self-bred modern varieties and foreign-introduced varieties. As modern varieties replaced the old, the gap between these indicators and foreign varieties gradually narrows.

Unlike the previous approach of only judging the seed size based on one-dimensional data, our study introduced 3D phenotypic indicators in addition to traditional one-dimensional data for seed classification, as opposed to solely assessing seed size. Our study revealed that between 1958-2020, seed morphology in China transitioned from first changing from large and small seed groups to three seed groups of large, medium and small, and finally to the change process of middle and small seed groups. The shift in seed size may be due to two primary factors. First, the medium seed group exhibits a more significant response to nitrogen use than both the larger and smaller seed groups. Large seed varieties exhibit longer fibers, greater fiber strength, improved uniformity in fiber length, and smaller particle sizes compared to small seed varieties (Main et al., 2014). Second, in terms of nutrient effects, small seeds exhibit faster germination and emergence compared to large seeds. Crops with larger seeds necessitate greater nutrient accumulation for germination and emergence, potentially impacting seed health and uniform emergence rates (Wang et al., 2008; Vidak et al., 2022). Consequently, breeders tend to favor smaller seeds in the cultivation process to achieve optimal growth conditions.

4.3 Seed morphological structure and germination prediction

Researchers often use micro-CT to analyze the structural characteristics of seeds and predict their germination potential. Through micro-CT images of chili peppers and germination tests, Ahmed et al. (2020) found that the shape and length of the embryonic root correlated with the germination quality of seeds. They believed that seeds with compact internal morphological structures (low air cavity ratio) and appropriate kernel shapes were able to germinate better. Additional investigation is required to further explore the correlation between the volume and proportion of the morphological structure of cotton seeds and the process of cotton seed germination. However, some studies have shown that certain doses of X-rays may kill or cause mutations in seeds, leading to abnormal kernel morphological structures such as degeneration, folding, lateral bending, and fracture (Gargiulo et al., 2020). Therefore, when exploring the relationship between seed morphology and germination, attention should be paid to the effects of X-ray radiation dose and exposure time on seed germination.

5 Conclusions

Using micro-CT scanning, this study conducted quantitative and comparative analyses of the morphological structure of 102 cotton seeds through 3D reconstruction and image segmentation. A non-destructive high-throughput analysis method was established to accurately identify the linear size, volume, and other indicators of seeds, as well as quantify phenotype indicators such as seed surface area, seed kernel volume, seed kernel surface area, cavity volume, seed coat volume, and average seed coat thickness. The study demonstrated a positive correlation between seed kernel size and seed size, while seed cavity size and average seed thickness were less influenced by other morphological indicators. During the period between 1904 and 2020, the overall trend in the physical morphological structure of cotton seeds in China decreased. However, for locally-bred cotton varieties (1958-2020), the size of the physical morphological structure of the seed increased, then decreased, demonstrating an overall increasing trend in size. Cluster analysis results showed that the seed type of China's independently bred cotton varieties underwent a transformation from large and small seed groups to large, medium, and small seed groups, and then to medium and small seed groups. The study proposes three seed morphological structure evaluation indicators, indicating that with the replacement of varieties, the specific surface area of the seed increases, the relative thickness of the seed coat increases, and the internal morphological structure of the seed becomes denser. Overall, these findings demonstrate that the morphological evolution history of cotton seeds in China provides important theoretical support for cotton variety breeding and seed quality evaluation.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

XG conceived and designed the experiments. YL, DL, and MG carried out the experiment and collected image data. YL and GH performed data processing and analysis, and drafted the manuscript. XG and YoZ is the supervisor of YL and participated in the guiding the writing of this manuscript. XL, SG, and YiZ provided the technical guidance and editing support. All authors contributed to the article and approved the submitted version.

Funding

This work was partially supported by Construction of Collaborative Innovation Center of Beijing Academy of Agricultural and Forestry Sciences (KJCX201917) and the earmarked fund for CARS-054.

Acknowledgments

We gratefully appreciate Professor Yanying Qu, from the Engineering Research Center of the Cotton Ministry of Education/College of Agronomy, Xinjiang Agricultural University; and Professor Keru Wang, from the Institute of Crop Sciences, Chinese Academy of Agricultural Science/Key Laboratory of Crop Physiology and Ecology, Ministry of Agriculture, for providing the cotton.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1219476/full#supplementary-material

References

Ahmed, M. R., Yasmin, J., Wakholi, C., Mukasa, P., and Cho, B. (2020). Classification of pepper seed quality based on internal structure using X-ray CT imaging. *Comput. Electron. Agric.* 179, 105839. doi: 10.1016/j.compag.2020.105839

Bernard, A., Hamdy, S., Le Corre, L., Dirlewanger, E., and Lheureux, F. (2020). 3D characterization of walnut morphological traits using X-ray computed tomography. *Plant Methods* 16 (1), 115. doi: 10.1186/s13007-020-00657-7

Bhawana, Miller, J. L., and Cahoon, A. B. (2014).3D Plant cell architecture of Arabidopsis thaliana (Brassicaceae) using focused ionz beam-scanning electron microscopy. *App. Plant Sci.App. Plant Sci.* 2 (6), 1300090. doi: 10.3732/apps.1300090

Brodersen, K. E., and Kuhl, M. (2023). Photosynthetic capacity in seagrass seeds and early-stage seedlings of Zostera marina. *New Phylogoist* 239 (4), 1300–1314. doi: 10.1111/nph.18986

Craig, S., Goodchild, D. J., and Miller, C. (1979).Structural aspects of protein accumulation in developing pea cotyledons. II* Three-dimensional reconstructions of vacuoles and protein bodies from serial sections. *Australian J. Plant Physiol.* 6, 81–98. doi: 10.1071/PP9820689

Crozier, D., Riera-Lizarazu, O., and Rooney, W. L. (2022). Application of X-ray computed tomography to analyze the structure of sorghum grain. *Plant Methods* 18, 3. doi: 10.1186/s13007-022-00837-7

Dong, P. F., Xie, R. Z., Wang, K. R., Ming, B., Hou, P., Hou, J. F., et al. (2020). Kernel crack characteristics for X-ray computed microtomography (μCT) and their relationship with the breakage rate of maize varieties. *J. Integr. Agric.* 19 (11), 2680–2689. doi: 10.1016/S2095-3119(20)63230-0

Fanuel, M., Ropartz, D., Guillon, F., Saulnier, L., and Rogniaux, H. (2018). Distribution of cell wall hemicelluloses in the wheat grain endosperm: a 3D perspective. *Planta* 248 (6), 1505–1513. doi: 10.1007/s00425-018-2980-0

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., et al. (2012). 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging* 30 (9), 1323–1341. doi: 10.1016/j.mri.2012.05.001

Feng, P. Y., Song, Y. L., and Wang, C. H. (2008). Causes and preventive measures for reduced germination rate during maize seed production. *Crops* 4, 105–107. doi: 10.3969/j.issn.1001-7283.2008.04.038

Gargiulo, L., Leonarduzzi, C., and Mele, G. (2020). Micro-CT imaging of tomato seeds: Predictive potential of 3D morphometry on germination. *Biosyst. Eng.* 200, 112–122. doi: 10.1016/j.biosystemseng.2020.09.003

Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., et al. (2010). A genetic framework for grain size and shape variation in wheat. *Plant Cell* 22 (4), 1046–1056. doi: 10.1105/tpc.110.074153

Gubatz, S., Dercksen, V. J., Brüß, C., Weschke, W., and Wobus, U. (2007). Analysis of barley (Hordeum vulgare) grain development using three-dimensional digital models. *Plant J.* 52 (4), 779–790. doi: 10.1111/j.1365-313X.2007.03260.x

Guelpa, A., Plessis, A. D., Kidd, M., and Manley, M. (2015). Non-destructive estimation of maize (Zea mays L.) kernel hardness by means of an X-ray micro-computed tomography (μCT) density calibration. *Food Bioprocess Technol.* 8 (7), 1419–1492. doi: 10.1007/s11947-015-1502-3

Hacisalihoglu, G., and Armstrong, P. (2023). Crop seed phenomics: focus on non-destructive functional trait phenotyping methods and applications. *Plants* 12, 1177. doi: 10.3390/plants12051177

Hao, Y. Y., Liu, Y., Chen, Y. Z., Han, L., Peng, J. C., Tang, S. J., et al. (2022). EISeg: an efficient interactive segmentation tool based on paddlepaddle. *arXiv* 2022, 2210. doi: 10.48550/arXiv.2210.08788

Hou, J. F., Zhang, Y., Jin, X. L., Dong, P. F., Guo, Y. N., Wang, K. R., et al. (2019). Structural parameters for X-ray micro-computed tomography (μCT) and their relationship with the breakage rate of maize varieties. *Plant Methods* 15, 161. doi: 10.1186/s13007-019-0538-1

Hu, M. J., Zhou, Y., Tang, Z. C., Xu, Z. Y., Wang, H. L., and Wang, P. T. (2018). Mechanical properties of the delinted and coated cottonseed. *J. Anhui Agric. Univ.* 45 (1), 175–180. doi: 10.13610/j.cnki.1672-352x.20180302.010

Hu, W. J., Zhang, C., Jiang, Y. Q., Huang, C. L., Liu, Q., Xiong, L. Z., et al. (2020). Nondestructive 3D image analysis pipeline to extract rice grain traits using X-Ray computed tomography. *Plant Phenomics* 3, 1–12. doi: 10.34133/2020/3414926

Huang, L., Yang, S. X., Wu, L. Y., Xin, Y., Song, J. K., Wang, L., et al. (2022). Genome-wide analysis of the *GW2*-like genes in *Gossypium* and functional characterization of the seed size effect of *GhGW2-2D*. *Front. Plant Sci.* 13, 860922. doi: 10.3389/fpls.2022.860922

Kranner, I., Kastberger, G., Hartbauer, M., and Pritchard, H. W. (2010). Noninvasive diagnosis of seed viability using infrared thermography. *Pnas* 107 (8), 3912–3917. doi: 10.1073/pnas.0914197107

Lang, J., Fu, Y. X., Zhou, Y., Cheng, M. P., Deng, M., and Li, M. L. (2021). Myb10-D confers PHS-3D resistance to pre-harvest sprouting by regulating *NCED* in ABA biosynthesis pathway of wheat. *New Phytol.* 230 (5), 1940–1952. doi: 10.1111/nph.17312

Le, T. D. Q., Alvarado, C., Girousse, C., Legland, D., and Chateigner-Boutin, A. (2019). Use of X-ray micro computed tomography imaging to analyze the morphology of wheat grain through its development. *Plant Methods* 15, 84. doi: 10.1186/s13007-019-0468-y

Liu, H. J., Bruning, B., Garnett, T., and Berger, B. (2020). Hyperspectral imaging and 3D technologies for plant phenotyping: From satellite to close-range sensing. *Comput. Electron. Agric.* 175, 105621. doi: 10.1016/j.compag.2020.105621

Liu, X. L., Yan, J., Li, Y. H., Pan, Z. X., and Fan, L. Q. (1997). Influences of seed quality on growth and yield in cotton. *J. Shanxi Agric. Sci.* 25 (4), 40–43.

Lorensen, W. E., and Cline, H. E. (1987). Marching cubes: a high resolution 3D surface construction algorithm. *Seminal Graphics* 21 (4), 163–169. doi: 10.1145/37402.37422

Lusic, H., and Grinstaf, M. W. (2013). X-ray-computed tomography contrast agents. *Chem. Rev.* 113 (3), 1641–1666. doi: 10.1021/cr200358s

Maeda, A. B., Wells, L. W., Sheehan, M. A., and Dever, J. K. (2021). Stories from the greenhouse—a brief on cotton seed germination. *Plants* 10 (12), 2807. doi: 10.3390/plants10122807

Main, C. L., Barber, L. T., Boman, R. K., Chapman, K., Dodds, D. M., and Duncan, S. (2014). Effects of nitrogen and planting seed size on cotton growth, development, and yield. *Agron. J.* 105 (6), 1853–1859. doi: 10.2134/agronj2013.0154

Metzner, R., Eggert, A., van Dusschoten, D., Pfugfelder, D., Gerth, S., Schurr, U., et al. (2015). Direct comparison of MRI and X-ray CT technologies for 3D imaging of root systems in soil: potential and challenges for root trait quantification. *Plant Methods* 11 (1), 17. doi: 10.1186/s13007-015-0060-z

Mohoric, A., Vergeldt, F. J., Gerkema, E., van Dalen, G., van den Doel, L. R., van Vliet, L. J., et al. (2009). The effect of rice kernel microstructure on cooking behaviour: A combined μ-CT and MRI study. *Food Chem.* 115 (4), 1491–1499. doi: 10.1016/j.foodchem.2009.01.089

Neubert, P., and Protzei, P. (2014). "Compact watershed and preemptive SLIC: on improving trade-offs of superpixel segmentation algorithms," in *2014 22nd International Conference on Pattern Recognition*, Stockholm, Sweden, 2014. 996–1001.

Ogawa, Y., Sugiyama, J., Kuensting, H., Ohtain, T., Hagiwara, S., Kokubo, M., et al. (2000). Development of visualization technique for three-dimensional distribution of protein and starch in a brown rice grain using sequential stained sections. *Food Sci. Technol. Res.* 6 (3), 176–178. doi: 10.3136/fstr.6.176

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernetics* 9 (1), 62–66. doi: 10.1109/TSMC.1979.4310076

Ozarslan, C. (2002). Physical properties of cotton seed. *Biosyst. Eng.* 83 (2), 169–174. doi: 10.1006/bioe.2002.0105

Pahlavani, M. H., and Abolhasani, K. (2006). Xenia effect on seed and embryo size in cotton (*Gossypium hirsutum* L.). *J. Appl. Genet.* 47 (4), 331–335. doi: 10.1007/BF03194642

Prior, D. A. M., Ooarka, K. J., and Roberts, I. M. (1999). En bloc optical sectioning of resin-embedded specimens using a confocal laser scanning microscope. *J. Microscopy* 193, 20–27. doi: 10.1046/j.1365-2818.1999.00433.x

Rashid, N., Zafar, M., Ahmad, M., Memon, R. A., Akhter, M. S., Malik, K., et al. (2020). Seed morphology: An addition to the taxonomy of Astragaleae and Trifolieae (Leguminosae: Papilionoidae) from Pakistan. *Microscopy Res. Technique* 84 (5), 1053–1062. doi: 10.1002/jemt.23666

Ruan, Y. L. (2005). Recent advances in understanding cotton fiber and seed development. *Seed Sci. Res.* 15 (4), 269–280. doi: 10.1079/SSR2005217

Schüler, P., Speck, T., Bührig-Polaczek, A., and Fleck, C. (2014). Structure-function relationships in Macadamia integrifolia seed coats - fundamentals of the hierarchical microstructure. *PloS One* 9 (8), e102913. doi: 10.1371/journal.pone.0102913

Su, Y., and Xiao, L. T. (2020). 3D visualization and volume-based quantification of rice chalkiness in *vivo* by using high resolution micro-CT. *Rice* 13, 69. doi: 10.1186/s12284-020-00429-w

Vidak, M., Lazarević, B., Javornik, T., Šatović, Z., and Carović-Stanko, K. (2022). Seed water absorption, germination, emergence and seedling phenotypic characterization of the common bean landraces differing in seed size and color. *Seeds* 1 (4), 324–339. doi: 10.3390/seeds1040027

Wang, Y. (2007). Identifying essentiality and methods of cotton seed vitality. *Seed Sci.&Technol.* 25 (3), 50. doi: 10.3969/j.issn.1005-2690.2007.03.031

Wang, X. Y., Lei, B. S., and Li, Y. Q. (2008). Effect of seed size on the plant growth at seedling stage in Cotton. *J. Anhui Agric. Sci.* 36 (23), 9939–9940. doi: 10.13989/j.cnki.0517-6611.2008.23.082

Warning, A., Verboven, P., Nicolaï, B., Van Dalen, G., and Datta, A. K. (2014). Computation of mass transport properties of apple and rice from X-ray microtomography images. *Innovative Food Sci. Emerging Technol.* 24 (8), 14–27. doi: 10.1016/j.ifset.2013.12.017

Wu, L. Y., Jia, B., Pei, W. F., Wang, L., Ma, J. J., Wu, M., et al. (2022). Quantitative trait locus analysis and identification of candidate genes affecting seed size and shape in an interspecific backcross inbred line population of *Gossypium hirsutum* × *Gossypium barbadense*. *Front. Plant Sci.* 13, 837984. doi: 10.3389/fpls.2022.837984

Wu, D., Wu, D., Feng, H., Duan, L. F., Dai, G. X., Liu, X., et al. (2021). A deep learning-integrated micro-CT image analysis pipeline for quantifying rice lodging resistance-related traits. *Plant Commun.* 2 (2), 100165. doi: 10.1016/j.xplc.2021.100165

Yamane, K., Oi, T., and Taniguchi, M. (2022). Evaluation of the validity of large-scale serial sectioning TEM for three-dimensional reconstruction of rice mesophyll cells and chloroplasts. *Protoplasma* 259 (5), 1219–1231. doi: 10.1007/s00709-021-01728-9

Yin, X. B., Hou, J. F., Ming, B., Zhang, Y., Guo, X. Y., and Gao, S. (2021). Kernel position effects of grain morphological characteristics by X-ray micro-computed tomography (μCT). *Int. J. Agric. Biol. Eng.* 14 (2), 159–166. doi: 10.25165/j.ijabe.20211402.6039

Yu, S. X. (2018). The development of cotton production in the recent hundred years of China. *J. Agric.* 8 (1), 85–91. doi: 10.11923/j.issn.2095-4050.cjas2018-1-093

Yu, L. J., Liu, L. B., Yang, W. N., Dan, W., Wang, J. H., He, Q., et al. (2022). A non-destructive coconut fruit and seed traits extraction method based on Micro-CT and deeplabV3+model. *Front. Plant Sci.* 13, 1069849. doi: 10.3389/fpls.2022.1069849

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., et al. (2006). User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31 (3), 1116–1128. doi: 10.1016/j.neuroimage.2006.01.015

Zhong, C. X., Chen, W., Liu, T., Hao, X. N., Li, Z. M., and Sun, C. M. (2016). Three-dimensional information measurement of wheat grain based on image technology. *Guangdong Agric. Sci.* 43 (2), 150–155. doi: 10.3969/j.issn.1004-874X.2016.02.028

Zhou, X., Zhao, C. J., Sun, J., Cao, Y., Yao, K. S., and Xu, M. (2023). A deep learning method for predicting lead content in oilseed rape leaves using fluorescence hyperspectral imaging. *Food Chem.* 409, 135251. doi: 10.1016/j.foodchem.2022.135251

Zong, Y., Yao, S., Crawford, G. W., Fang, H., Lang, J. F., Fan, J. D., et al. (2017). Selection for oil content during soybean domestication revealed by X-ray tomography of ancient beans. *Sci. Rep.* 7, 43595. doi: 10.1038/srep43595

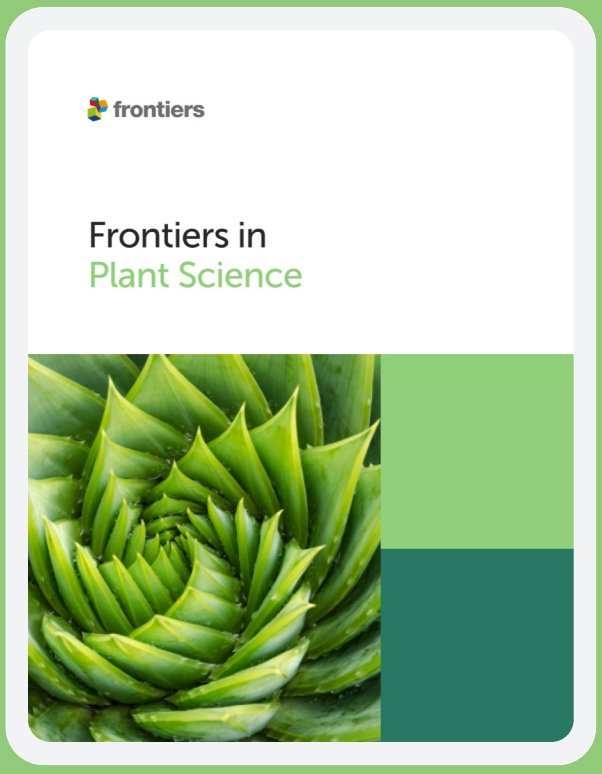
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact