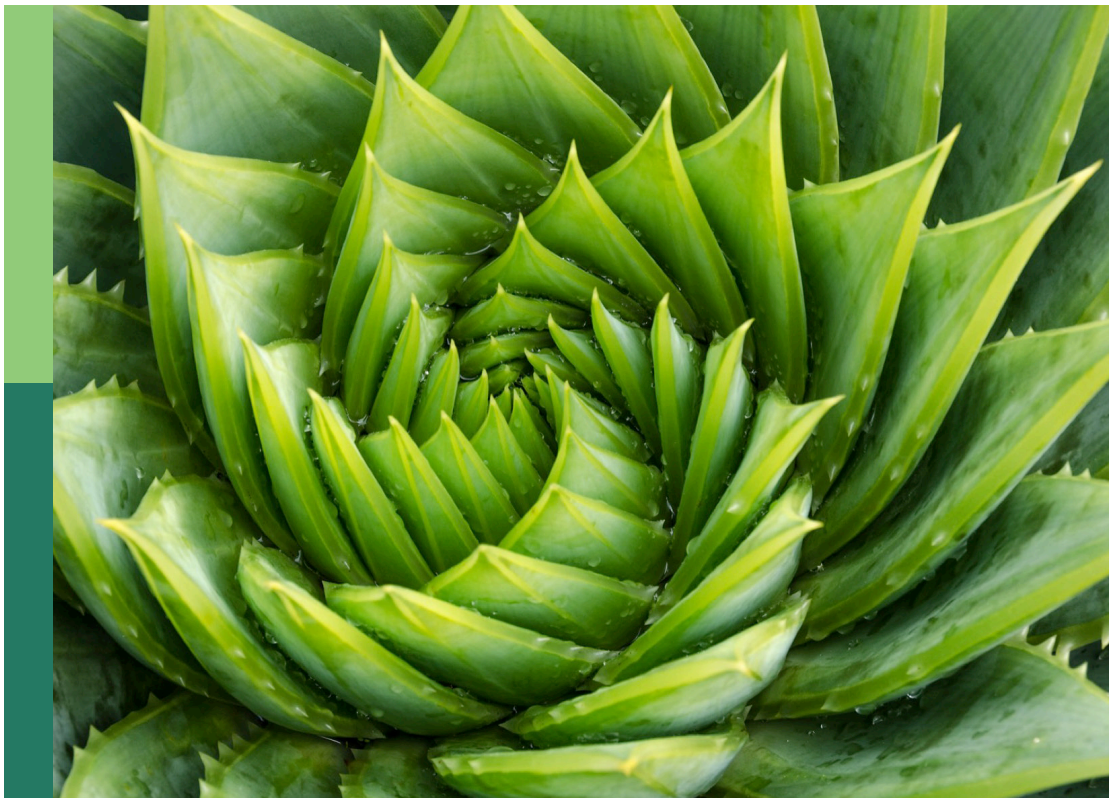
frontiers | Research Topics

AI, sensors and robotics in plant phenotyping and precision agriculture

Edited by
Yongliang Qiao, João Valente, Zhao Zhang, Daobilige Su and Dongjian He

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-83250-977-7
DOI 10.3389/978-2-83250-977-7

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

AI, sensors and robotics in plant phenotyping and precision agriculture

Topic editors

Yongliang Qiao — The University of Sydney, Australia
João Valente — Wageningen University and Research, Netherlands
Zhao Zhang — China Agricultural University, China
Daobilige Su — China Agricultural University, China
Dongjian He — Northwest A&F University, China

Citation

Qiao, Y., Valente, J., Zhang, Z., Su, D., He, D., eds. (2022). *AI, sensors and robotics in plant phenotyping and precision agriculture*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-977-7

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 23 November 2022
DOI 10.3389/fpls.2022.1064219

OPEN ACCESS

EDITED AND REVIEWED BY
Chuanlei Zhang,
Tianjin University of Science and Technology, China

*CORRESPONDENCE
Daobilige Su
sudao@cau.edu.cn

SPECIALTY SECTION
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

RECEIVED 08 October 2022
ACCEPTED 04 November 2022
PUBLISHED 23 November 2022

CITATION
Qiao Y, Valente J, Su D, Zhang Z and He D (2022) Editorial: AI, sensors and robotics in plant phenotyping and precision agriculture.
Front. Plant Sci. 13:1064219.
doi: 10.3389/fpls.2022.1064219

Editorial: AI, sensors and robotics in plant phenotyping and precision agriculture

Yongliang Qiao[1], João Valente[2], Daobilige Su[3]*, Zhao Zhang[4,5] and Dongjian He[6]

[1]Australian Centre for Field Robotics (ACFR), Faculty of Engineering, The University of Sydney, Sydney, NSW, Australia, [2]Information Technology Group, Wageningen University & Research, Wageningen, Netherlands, [3]College of Engineering, China Agricultural University, Beijing, China, [4]Key Laboratory of Smart Agriculture System Integration, Ministry of Education, China Agricultural University, Beijing, China, [5]Key Laboratory of Agriculture Information Acquisition Technology, Ministry of Agriculture and Rural Affairs of China, China Agricultural University, Beijing, China, [6]College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi, China

KEYWORDS

artificial intelligence, plant phenotyping, precision agriculture, smart sensors, agricultural robotics, UAV

Editorial on the Research Topic
AI, sensors and robotics in plant phenotyping and precision agriculture

Introduction

Plants and their production play an important role in retaining the sustainability for the natural ecosystem and human beings' food security. With the increasing global population, rapid urbanization and climate change, how to improve plant protection levels, increase plant breeding speed and make sure the agricultural planting in a sustainable and low-carbon dioxide manner becomes challenging. One way to address this issue is to develop the technology of plant phenotyping and precision agriculture (Costa et al.). Plant phenotyping and precision agriculture as information- and technology-based approaches, could evaluate a large amount of plants and provide effective information to production management. Plant phenotyping assesses complex plant traits such as plant morphology, plant stress, crop yield, plant physiological, anatomical traits, and genotype performance under distinct environmental conditions. Precision agriculture is aimed at examining spatial heterogeneities within crop stands based on the spatial and temporal variability in crop and soil factors within a field (Stafford, 2000 and Patrício and Rieder, 2018). High-throughput phenotyping in precision agriculture is helpful to improve management practices, and efficient phenotyping in the field also reduces the invested resources (e.g., fertilizer, water, pesticide).

In recent years, cutting-edge technologies for plant phenotyping and precision agriculture are fundamental to improve the productivity and sustainability of plant production systems (Narvaez et al., 2017). Especially, the development of Artificial Intelligence (AI), smart sensors and robotics provides a non-invasive manner for assessing complex traits in plants (as shown in Figure 1), measuring plant-physiological parameters, diagnosing plant diseases, predicting the yield and performance of plants at various organizational scales (Purcell and Neubauer, 2023).

The comprehensive plant phenotyping emerges from the dynamic and local interaction of phenotypes with the spatially and temporally dynamic environment above and below ground, while assessing complex plant traits such as growth, tolerance, resistance, physiology, ecology, plant stress and yield, which benefits the farmers and plant breeders to identify phenotyping parameters and select desirable genotypes that provide effective information to make informed agricultural production management decisions (Li et al., 2014). By assessing complex plant traits (e.g., growth, development, resistance, physiology, ecology), high yielding and stress-tolerance crop varieties adapt to future climate conditions and resistant to pests and diseases, produce enough food, feed, fiber, and fine chemicals in next century to meet the needs of a growing population worldwide (Abbasi et al., 2022).

Plant phenotyping

Plant monitoring and phenotyping can reflect many valuable parameters and effective information for optimizing agricultural production management in smart farming. Traditional manual based methods rely on experienced farmers, which is of low-accuracy and poor efficiency. Nowadays, a range of sensors (various RGB, multi-and hyperspectral cameras, 3D-sensors, etc.) and platforms have been used to realize real-time, rapid, and efficient plant phenotyping. According to different perception principles, these sensors mainly have ground feature spectrometers, spectral imaging sensors, and other imaging spectrometers.

Qin et al. proposed a real-time and low-cost Ag-YOLO model for crop monitoring and crop spraying, which achieved 0.92 F1-score with a speed of 36.5 frames per second (fps) on Intel Neural Compute Stick 2 (NCS2). Liu et al. proposed a portable wild phenotyping system based on segmentation results from DeepLabV3+ model to obtain 45 traits, including 15 plant traits, 25 leaf traits and 5 stem traits. The proposed system provides a solution for maize phenotyping in the field and benefit crop breeding. Lu et al. proposed a soybean yield in-field prediction method based on bean pods and leaves image recognition using a deep learning algorithm combined with a generalized regression neural network (GRNN). According to the experiments, the soybean yield of each planter was obtained by accumulating the weight of all soybean pod types and the average accuracy was up to 97.43%.

In addition, an identification model YOLO-VOLO-LS was constructed for hydroponic lettuce grown in a greenhouse under the conditions of different growth periods (Zhang and Li). By combining the respective advantages of the target detection mechanism and the classification mechanism, a nearly 100% of the lettuce classification effect in the growth stages of days 1, 6, 12, 18, 24, and 30 were achieved. Wang et al. proposed a

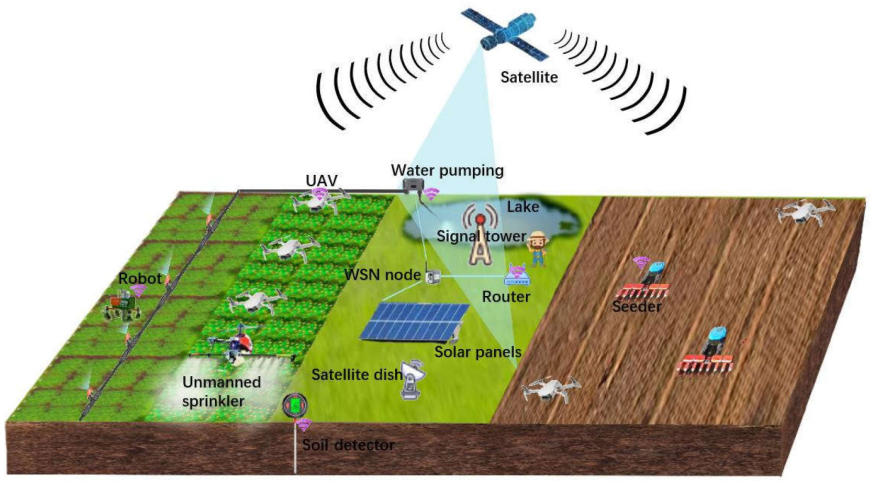


FIGURE 1
AI, sensors and robotics in plant phenotyping and precision agriculture.

lightweight model based on the improved You Only Look Once version 4 (YOLOv4) to detect dense plums in orchards, which achieved 86.34% detection accuracy.

For obtaining image-based phenotypic information of wheat traits for spike morphology analysis and yield estimation, Zhang et al. collected wheat images from fields and proposed an optimized hybrid task cascade model for automatic dense wheat spike segmentation. Experimental results showed that they achieved an average precision (AP) of 90.7%, and an accuracy of 99.29%.for wheat spike counting. Qiu et al. processed the color images of the spike in YCbCr color space and then utilized Faster R-CNN to detect the spikelets. Testing results showed that the root mean squared errors between the automatic and manual counted spikelets for four wheat lines were 0.62. Qi et al. proposed a novel tea chrysanthemum–generative adversarial network (TC-GAN) for tea chrysanthemum detection, which achieved an optimal average precision (AP) of 90.09%.

Nitrate nitrogen plays an important role during crop growth, and the operation of Increasing N fertilizer dosage and application is usually one of the essential ways to boost crop productivity. Su et al. proposed an ISE system combined with a temperature sensor and a pH electrode to automatically measure the concentrations of Nitrate nitrogen.

As sugar being the energy source of plants and plays an important role in plant growth and development, Liu et al. developed an enzyme-free electrochemical sensor for *in situ* detection of reducing sugar, which demonstrated that the COOH-GR–COOH-MWNT–AuNP-modified electrode exhibited a good catalysis behavior. To investigate the study effect of vegetation distribution on mean flow velocity and turbulence characteristics in a channel, Wang et al. constructed a flow velocity distribution model to study the microscopic mechanism of the flow velocity distribution in the upper layer of vegetation, which provides a solution for flow measurement in the ecological channel.

Hyperspectral imaging is advantageous in delivering reliable and comprehensive analysis of characteristics or properties of plant materials, which is a powerful modality for measuring spectral and spatial information of samples simultaneously (Lu et al., 2020). Lu et al. classified industrial hemp cultivars, growth stages, and plant organs (leaves vs. flowers) using hyperspectral imaging technology. Based on regularized linear discriminant analysis, an accuracy of up to 99.6% was achieved in differentiating the five hemp cultivars. Liu et al. designed a near-infrared (NIR) phenotypic sensor for predicting wheat gain quality, and the R2 of the relative diffuse reflectance (RDR) of all four wavelengths of the phenotypic sensor and the reflectance of the diffusion fabrics were higher than 0.99.

Plant diseases detection

Diseases are the main causal factors affecting crop growth and yield. Reliable and timely plant disease detection is important for plant protection activities, field crop growth and plant breeding. AI and computer vision based diagnosis and detection of plant diseases must consider that the occurrence of plant disease depends on specific environmental factors and diseases often exhibit a heterogeneous distribution in fields (Mahlein, 2016).

Wang et al. proposed a YOLOv3-tiny-IRB algorithm to improve the detection accuracy of tomato diseases and pests under conditions of occlusion and overlapping in real natural environment, which achieved the mean average precision (mAP) of 98.3, 92.1, and 90.2%, respectively under three conditions: (a) deep separation, (b) debris occlusion, and (c) leaves overlapping. Zhang et al. proposed YOLOv5-CA based GDM detection approach for grape downy mildew disease detection, and the experimental results show that the proposed YOLOv5-CA achieved a detection precision of 85.59%, a recall of 83.70%, and a mAP@0.5 of 89.55%, which are superior to the popular methods, including Faster R-CNN, YOLOv3, and YOLOv5.

Chen et al. proposed 2D histogram Otsu based approach for segmenting maize foliar disease images, the experimental results indicated that the method effectively improved the segmentation of the three maize disease spot images and could obtain more apparent disease spot areas. Zhang et al. extracted handcrafted and deep features from the color image and color-infrared (CIR) image, and the DFs coupled with parallel feature fusion resulted in diagnosis accuracies of over 70%.

Robotics and UAVs in smart farming

Robotics and UAVs have shown great efficiency and effectiveness in the agriculture field. In recent years, many agricultural related robotics and UAVs have been designed and developed to manage crops, plants, livestock and fishes (Qiao et al., 2019; Su et al., 2021; Li et al., 2022 and Du et al., 2022). Based on the Simultaneous Localization and Mapping (SLAM), place recognition and autonomous navigation, robots or UAVs can autonomously drive and perform actions such as harvesting, picking and trimming.

In robotic precision spray of vegetables, accurate and reliable detection and tracking of every vegetable is of utmost importance. Hu et al. proposed LettuceMOT, a multiple object tracking (MOT) method to correlate these re-appeared vegetables with their previous identities. The experimental results show that LettuceMOT outperformed existing state-of-the-art MOT methods (e.g., ByteTrack, FairMOT, TraDeS and SORT).

To achieve the rapid harvesting of table grapes planted with a standard trellis in the grape industry, Jiang et al. carried out a dual-arm high-speed grape-harvesting robot to improve low picking efficiency. Robotic arm and camera view analysis of the workspace harvesting process was performed using MATLAB, and it can be concluded that the structural design

of this robot meets the grape harvesting requirements with a standard trellis. The field performance test verifies that the average harvesting cycle of the robot with both arms reached 9 s/bunch, and the success rate of bunch identification and harvesting success rate reached 88% and 83%, respectively, which were significantly better than those of existing harvesting robots worldwide.

In terms of agricultural navigation technologies, Xie et al. proposed the miniaturization scheme of zooming detection arc based on variable central angle and established the adjustment equation of the detection distance of photoelectric switches at each position, a small integrated photoelectric arc array navigation sensor with a cost of about $65 is developed using an embedded microcontroller. However, there is still a problem of external noise and other factors causing the failure of the navigation system. To solve this problem, Lv et al. proposed an agricultural scene-based multi-sensor fusion method *via* a loosely coupled extended Kalman filter algorithm to reduce interference from external environment. Specifically, the proposed method fuses inertial measurement unit (IMU), robot odometer (ODOM), global navigation and positioning system (GPS), and visual inertial odometry (VIO), and uses visualization tools to simulate and analyze the robot trajectory and error. In experiments, the high accuracy and the robustness of the proposed algorithm were verified when sensors fail. The experimental results show that the proposed algorithm has better accuracy and robustness on the agricultural dataset than other algorithms.

For phenotypic feature detection in the study of automatic trimming, Tang et al. optimized and designed a long-belt finger-clip precision seed metering device, which includes a diffuse reflection photoelectric sensor and rectangular optical fiber sensor to monitor the number of corn seeds in the seeding process. To automatically trim hedges, Zhang et al. proposed a binocular vision-based shape reconstruction and measurement system, based on stereo correcting algorithm and an improved semi-global block matching (SGBM) algorithm The center coordinate and radius of the spherical hedges can be measured. The outdoor test shows that the average error and average relative error of spherical hedges radius by the proposed system are 4.02 mm and 0.44%, respectively. The average location deviation of the center coordinate of spherical hedges is 18.29 mm.

Fu et al. quantified the forces on the stalks caused by root anchorage in corn harvesting, and a root force measurement system was designed and applied in this study. The bending moment and torsional moment on the upright and lodged corn stalks were measured in corn harvesting. By analyzing the bending moment curves on the lodged corn stalks, it was proposed that for the harvesting of corn lodged in the forward, reverse, and lateral direction, the corresponding harvester header improvement suggestions are enlarging the size of pins on the gathering chains, reducing the speed of gathering chains, and lengthening the snouts with a sleeker surface, respectively. This study provides base data for the root anchorage effect on lodged corn and provides references for the improved design of the corn harvester header.

On the other hand, UAV based platform are also popular in the precision agriculture and plant phenotyping applications because their large cover range and higher data scanning speed. In addition, UAVs can fly automatically with less human intervention during data collection. Zhaosheng et al. improved wheat ears identification performance in a field environment using improved YOLOX-m model. To develop a data processing pipeline for performing fast and accurate pre-harvest yield predictions of cotton breeding fields using aerial image, Rodriguez-Sanchez et al. used a Support Vector Machine (SVM) classifier with four selected features to identify the cotton pixels present in each plot image, which achieved an accuracy of 89%, a precision of 86%, a recall of 75%, and an F1-score of 80% at recognizing cotton pixels. This study demonstrates that aerial imagery with machine learning techniques can be a reliable, efficient, and effective tool for pre-harvest cotton yield prediction. Krul et al., 2021 studied the feasibility to apply UAV for indoor farming monitoring and control. The performance of different state-of-the-art visual simultaneous localization and mapping (VSLAM) algorithms with a small and low-cost UAV was assessed. The authors found that ORB-SLAM was the algorithm that perform best in such an environment. Tests in the farming facilities where performed and different maps were generated.

Finally, agricultural management could also benefit from the collaboration between aerial and ground robotic systems. The aerial robotics could survey a field using different types of sensors and payloads. Moreover, it could provide the ground robot a detailed map with specific positions where the ground robot need to inspect further or perform some action with an actuator. Conesa-Muñoz et al., 2016 proposed a multi-robot system to reduce the amount of herbicide during site-specific treatments. The combination of aerial-ground robotics systems allows to reach a 97% spray accuracy and a mean deviation lower than 7cm. Zhang et al., 2022 investigated the spatial-variability of orchards flower blossom from an aerial and ground perspective. Several point clouds where acquired in a commercial orchard (Elstar) field using a UAV and a ground vehicles. The feasibility of combing data from both platforms to assess flowering intensity at the tree-level, was demonstrated, yielded R^2>0.7 and RMSE lower than 20.

Conclusions

Plant phenotyping and precision agriculture is becoming a very important topic for future agriculture. The increasing

population and climate change push us to take actions to plant crops against pests, diseases, and harsh environments (e.g. lack of nutrients, water, fertilizers or light). The new technologies such as AI, sensors and robotics enables farmers to take a data-driven approach to collect and analyze data to monitor the real-time status of the plans and crops to improve production yield quality. For precision agriculture, the grand challenges lie in identification of cheap, robust, easy-to-use, rapid and automated phenotyping methods that can feed into Decision Support System. In addition, the field environment will provide challenges in sometimes rapidly varying light conditions, wind and temperature, as well as combinations of multiple stresses. Despite all these challenges, automated and systematic stress detection by field-phenotyping holds great promise to accelerate Integrated pest management where on-farm live monitoring of stress and disease are key factors to reduce the reliance on pesticides.

In the future, the integration of automated data collection and analysis, AI algorithms, robotics and decision support systems will bring unmanned farming to our lives. Moreover, the ground-level or aerial-level robotic systems will also have a major role in plant phenotyping and precision agriculture, for monitoring, disease control and harvesting.

Author contributions

YQ: investigation, methodology, writing-original draft, and editing. JV, DS, ZZ, and DH: writing-review and editing. 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

References

Abbasi, R., Martinez, P., and Ahmad, R. (2022). The digitization of agricultural industry – a systematic literature review on agriculture 4.0. *Smart. Agric. Technol.* 2, 100042. doi: 10.1016/j.atech.2022.100042

Conesa-Muñoz, J., Valente, J., Del Cerro, J., Barrientos, A., and Ribeiro, A. (2016). A multi-robot sense-act approach to lead to a proper acting in environmental incidents. *Sensors* 16 (8), 1269. doi: 10.3390/s16081269

Du, L., Lu, Z., and Li, D. (2022). Broodstock breeding behaviour recognition based on Resnet50-LSTM with CBAM attention mechanism. *Comput. Electron. Agric.* 202, 107404. doi: 10.1016/j.compag.2022.107404

Krul, S., Pantos, C., Frangulea, M., and Valente, J. (2021). Visual SLAM for indoor livestock and farming using a small drone with a monocular camera: A feasibility study. *Drones* 5 (2), 41. doi: 10.3390/drones5020041

Li, J., Qiao, Y., Liu, S., Zhang, J., Yang, Z., and Wang, M. (2022). An improved YOLOv5-based vegetable disease detection method. *Comput. Electron. Agric.* 202, 107345. doi: 10.1016/j.compag.2022.107345

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14 (11), 20078–20111. doi: 10.3390/s141120078

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100 (2), 241–251. doi: 10.1094/PDIS-03-15-0340-FE

Narvaez, F. Y., Reina, G., Torres-Torriti, M., Kantor, G., and Auat Cheein, F. (2017). A survey of ranging and imaging techniques for precision agriculture phenotyping. *EEE/ASME. Trans. Mechatronics.* 22 (6), 2428–2439. doi: 10.1109/TMECH.2017.2760866

Patrício, 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

Purcell, W., and Neubauer, T. (2023). Digital twins in agriculture: A state-of-the-art review. *Smart. Agric. Technol.* 3, 100094. doi: 10.1016/j.atech.2022.100094

Qiao, Y., Truman, M., and Sukkarieh, S. (2019). Cattle segmentation and contour extraction based on mask r-CNN for precision livestock farming. *Comput. Electron. Agric.* 165, 104958. doi: 10.1016/j.compag.2019.104958

Stafford, J. V. (2000). Implementing precision agriculture in the 21st century. *J. Agric. Eng. Res.* 76 (3), 267–275. doi: 10.1006/jaer.2000.0577

Su, D., Qiao, Y., Kong, H., and Sukkarieh, S. (2021). Real time detection of inter-row ryegrass in wheat farms using deep learning. *Biosyst. Eng.* 204, 198–211. doi: 10.1016/j.biosystemseng.2021.01.019

Zhang, C., Mouton, C., Valente, J., Kooistra, L., van Ooteghem, R., de Hoog, D., et al. (2022). Automatic flower cluster estimation in apple orchards using aerial and ground based point clouds. *Biosyst. Eng.* 221, 164–180. doi: 10.1016/j.biosystemseng.2022.05.004

frontiers in Plant Science

ORIGINAL RESEARCH
published: 22 November 2021
doi: 10.3389/fpls.2021.753613

Study on Flow Velocity Distribution in Open Channel With Flexible Vegetation

*Shiyu Wang[1], Yi Zhou[1], Tongshu Li[1], Song Li[1], Mingwu Zhang[2] and Yu Han[1]**

[1] College of Water Resources and Civil Engineering, China Agricultural University, Beijing, China, [2] Yellow River Institute of Hydraulic Research, Yellow River Conservancy Commission, Zhengzhou, China

Ecological management of river channels is a hot topic for current sustainable development and flow measurement of ecological river is an important part. In this article, a flow velocity distribution model of the channel containing flexible vegetation is constructed from the vegetation riverbed theory and the bursting phenomenon to reveal the microscopic mechanism of the flow velocity distribution in the upper layer of vegetation. In the vegetation riverbed law, the effect of flexible vegetation is evaluated by the mixed length formula. The bursting phenomenon law considers the influence of the channel sidewalls on the flow and a two-dimensional velocity model is established by introducing the concept of average turbulence structure. The mechanism of the downward shift of the maximum flow velocity point on the channel sidewall is explained. The verification of the calculated velocity profiles is carried out based on data obtained in laboratory experiments. The results show that the combination of the two models can well describe the velocity distribution of the whole channel. At the end, the phenomenon of flow velocity zoning in open channel is discussed, which provides a solution for flow measurement in ecological channel.

Keywords: flexible vegetation, velocity distribution, open channel, bursting phenomenon, mixing length

OPEN ACCESS

Edited by:
Dongjian He,
Northwest A&F University, China

Reviewed by:
Muthusamy Ramakrishnan,
Nanjing Forestry University, China
Soumen Maji,
Central Institute of Technology Kokrajhar, India

***Correspondence:**
Yu Han
Yhan@cau.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *05 August 2021*
Accepted: *19 October 2021*
Published: *22 November 2021*

Citation:
Wang S, Zhou Y, Li T, Li S, Zhang M and Han Y (2021) Study on Flow Velocity Distribution in Open Channel With Flexible Vegetation. Front. Plant Sci. 12:753613. doi: 10.3389/fpls.2021.753613

INTRODUCTION

Sustainable development is an important issue in the world. The presence of vegetation in river systems contributes to the sustainability of rivers and enhances the self-cleaning capacity of water (Lozanovska et al., 2020). However, the existence of vegetation will change the channel resistance and raise the water level (Rivaes et al., 2017). Ecological discharge is an important parameter in the ecological channel. The existence of vegetation causes some interference on the flow field and flow monitoring. Therefore, it is necessary to study the velocity distribution of the channel with flexible vegetation.

The effect of vegetation distribution on mean flow velocity and turbulence characteristics in a channel is generally studied by model experiments. Järvelä (2002) used natural aquatic grass, sedge, and willows as test material to study the influence of plants on flow velocity distribution by introducing plant parameters that characterize the flexibility and morphology of plants in flow. Bennett et al. (2002) analyzed the influence of vegetation belts arranged at intervals between two sides of water channel on flow characteristics by model test. It was found that vegetation belts carried flow to the other side and the intensity of flow were related to the density of

vegetation. Wilson et al. (2005) developed a three-dimensional numerical model and compared it with a standard turbulent flow model to analyze the effect of willows on flow, pointing out that determining the degree of plant bending in flow is important for the study of the interaction mechanisms between vegetation and flow. Bouma et al. (2005) analyzed the influence of vegetation belt with specific shape on flow rate and turbulence and proposed that the distribution of flow rate is closely related to the spacing of vegetation belt. Neumeier (2005) measured the flow velocity and turbulence characteristics of the wake after it crossed the marsh vegetation and inferred that the larger vertical flow velocity and higher turbulence in front of the vegetation could reduce sediment deposition. Helmiö (2002) developed a one-dimensional model to solve the problem of composite channels with vegetation, in which the effect of flow distribution caused by vegetation was considered and the model was validated by using data from the Rhine river. He pointing out that vegetation density and the width of the diffuse zone have a large effect on the transport capacity of the main channel and significantly reduce the flood flow capacity of the channel. Leu et al. (2008) established a two-dimensional model to study the velocity distribution of five different submerged plants in different water depths, pointing out that the submerged plants can reduce the velocity and can meet the requirements of flood control. Zong and Nepf (2009) carried out model tests on flow in semi-covered vegetated river and discussed the changes of flow velocity and turbulence characteristics in vegetation and non-vegetation areas and found that flow can be divided into adjustment areas and fully developed areas. Siniscalchi et al. (2012) measured the flow characteristics of submerged flexible vegetation by flow channel experiment. It was found that the turbulence intensity corresponding to the front of vegetation and the top of vegetation increased and a negative Reynolds stress appeared in the vegetation layer, reflecting the influence of vegetation morphology on the flow velocity and turbulent flux. Wang et al. (2015) proposed a method to calculate the analytical solution of the vertical distribution of the mean flow velocity in the flow direction of an open channel containing flexible vegetation when a large bend occurs and derived a formula for the mean flow velocity of the linear flow resistance in the momentum equation. Okamoto et al. (2016) study how plant motions were coupled to strong oscillations in flow velocity associated with the "monami" phenomenon and its vertical extent in an open channel with flexible vegetation. Wang et al. (2016) simulated the water flow due to wind action on the water surface in a shallow lake through experiments. The effects of different densities of aquatic flexible vegetation on the hydrodynamic characteristics at different wind speeds were investigated. By analyzing the changes of vegetation drag (C_d) and friction (C_f) coefficient with flow and vegetation conditions under different test conditions, Pu et al. (2019) obtained a model of velocity distribution in open channel with flexible vegetation.

Although there are many studies on the hydrodynamic characteristics of flow with flexible vegetation, most of them are still limited to the corresponding hydrodynamic characteristics of specific vegetation. The characteristics of the flow field in the upper layer of vegetation have not been recognized after the flows through the vegetation and most of the current studies focused on the hydrodynamic characteristics of the flow in the center of the open channel. In fact, the hydrodynamic characteristics of the channel center and the open channel sidewalls are not the same. There are fewer studies on the phenomenon of downward shift of the maximum flow velocity point for the channel sidewall. The division of velocity zoning of channel section is not clear. The research results are not universal, so the research needs to be further strengthened.

This article studies on the velocity distribution law of ecological river with vegetation. Based on the vegetation riverbed theory and bursting phenomenon, the velocity distribution law of the whole section of the flow is established. Combined with the physical model test of ecological channel, the correctness of velocity distribution law is verified. Meanwhile, the flow velocity zoning phenomenon of the ecological open channel is discussed with the test results. It provides a solution for the flow measurement of ecological river.

MODELING THE VELOCITY DISTRIBUTION IN CHANNEL WITH FLEXIBLE VEGETATION

Velocity Distribution Based on the Vegetation Riverbed Theory

Righetti and Armanini (2002) proposed that the vegetation layer can be regarded as a large-scale roughness and it is considered as a part of the riverbed. It is called the vegetation riverbed theory and the non-vegetation layer flows on this vegetation riverbed (Huai et al., 2009).

In this article, we focus on the velocity distribution of the non-vegetation layer and the initial velocity of the non-vegetation layer is the velocity of the top layer of vegetation. Therefore, it is necessary to know the velocity of the vegetation top layer. For the vegetation layer, the corrected gravity term of the water volume occupied by the flexible vegetation can be ignored. For a given water volume, plant-induced resistance per fluid mass can be described as (Kubrak et al., 2008):

$$F = \frac{1}{2}\rho C_d m A u_{ud}^2 \tag{1}$$

Where, ρ is the density of water, m is the number of grass per unit area, A is the upstream area of vegetation, u_{ud} is the flow velocity at the top of the vegetation, and C_d is the drag force coefficient of the vegetation. For the drag coefficient of vegetation, the variation in vegetation layer from 1 to 1.5 was estimated from Kubrak et al. (2008) and Yang and Choi (2009) test data. It is found that Klopstra et al. (1997) proposed that it can be better applied to cylindrical vegetation. Therefore, this experiment is adopted C_d =1.4.

The shear stress of uniform flow in open channel can be expressed as:

$$\tau = \rho u_*^2 \tag{2}$$

Where, τ is the shear stress, u_* is the friction velocity and it can be expressed as (Tracy and Lester, 1961):

$$u_* = \sqrt{gHS} \tag{3}$$

Where, H is the depth of water and S is the slope of the channel. Since the vegetation layer is not a uniform flow, it is necessary to multiply a correction factor before Eq. (2), and combined with Eq. (1) can be obtained that:

$$\alpha \rho u_*^2 = \frac{1}{2}\rho C_d m A u_{ud}^2 \tag{4}$$

Where, α is the correction coefficient and different correction coefficients can be got for different flexible vegetation. In this experiment, we can take $\alpha = 7 \times \frac{h_v}{H}$, where h_v is the height of deflected vegetation. Eq. (4) can be simplified to get the expression of the top flow velocity of vegetation.

$$u_{ud} = \alpha \sqrt{\frac{2u_*^2}{C_d m A}} \tag{5}$$

For the non-vegetated layer, the equation of the flow under constant uniform flow conditions according to the force balance principle is:

$$\frac{\partial \tau}{\partial y} + \rho g S = 0 \tag{6}$$

Neglecting the viscous stress and integrating Eq. (6), the distribution relation of the tangential stress is obtained as:

$$\tau = \rho g S \left(H - y\right) \tag{7}$$

Prandtl (1925) assumes that the distance displaced by the momentum of the fluid mass before it is changed by the new environment is l. This is Prandtl's mixing-length theory. According to Prandtl's mixing-length theory, there are:

$$\frac{\tau}{\rho} = l^2 \left(\frac{\partial u}{\partial y}\right)^2 \tag{8}$$

According to the vegetation riverbed theory, the original channel is partly occupied by vegetation and the flow seems to be "compressed" and the expression of the mixing length should reflect this "compression" (Huai et al., 2009). A new mixed length expression is assumed as follows:

$$l = \left(\frac{H\text{-}h_v}{H}\right) \kappa y \sqrt{\frac{\tau}{\tau_{max}}} \tag{9}$$

According to Eq. (7), we get:

$$\tau_{max} = \rho g S H \tag{10}$$

Combining Eqs. (8–10), it is obtained that:

$$du = \frac{1}{\kappa} \frac{H}{H\text{-}h_v} \sqrt{gSH} \frac{1}{y} dy \tag{11}$$

The integral is organized to obtain:

$$\frac{u}{u_*} = \frac{1}{\kappa} \frac{H}{(H\text{-}h_v)} \ln\left(\frac{y}{h_v}\right) + \frac{u_{ud}}{u_*} \tag{12}$$

Equation (12) is the theoretical model of flow distribution in open channels containing vegetation based on the vegetation riverbed theory. This velocity equation is named as the vegetation riverbed law (VRL).

Velocity Distribution Based on the Bursting Phenomenon

As also mentioned earlier it can be proposed that the vegetation layer can be considered as a kind of rough object with a large scale, we can assume that the flexible vegetation is a special kind of boundary layer where turbulence can lead to the emergence of coherent structures. Nezu et al. (1993) introduced the concept of bursting phenomenon. The bursting phenomenon refers to the phenomenon that the boundary layer will suddenly rupture locally at a certain location in space when there is a strong interaction between the inner and outer regions of the boundary layer. In other words, the upthrow phenomenon of low-speed fluid and the down sweep phenomenon of high-speed fluid in wall turbulence are called bursting phenomenon. In fact, vortices originating from the wall region can be observed in the turbulent core and even at the free surface. This vortex is also intermittent and it is randomly distributed in size and direction (Yang, 2010).

For turbulent flows, the expressions for the mean Reynolds Navier–Stokes equations and the continuum equations are as follows:

$$u\left(\frac{\partial u}{\partial x}\right) + v\left(\frac{\partial u}{\partial y}\right) + w\left(\frac{\partial u}{\partial z}\right) = gS + \frac{\partial}{\partial x}\left(\overline{-u'u'}\right) + \frac{\partial}{\partial y}\left(\overline{-u'v'}\right) + \frac{\partial}{\partial z}\left(\overline{-u'w'}\right) + \mu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) \tag{13}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{14}$$

Where, u, v, and w represent the flow velocity along the flow direction x, y, and z. v is the kinematic viscosity, $\overline{-u'u'}$, $\overline{-u'v'}$, $\overline{-u'w'}$ is the turbulent shear stress. For uniform flow, the variation in x direction can be disregarded and Eq. (13) can be written as:

$$v\left(\frac{\partial u}{\partial y}\right) + w\left(\frac{\partial u}{\partial z}\right) = gS + \frac{\partial}{\partial y}\left(\overline{-u'v'}\right) + \frac{\partial}{\partial z}\left(\overline{-u'w'}\right) + \mu\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) \tag{15}$$

In 1883, Reynolds decomposed the instantaneous velocities into two contributions—mean velocities and velocity fluctuations. Equation (15) can be further written as:

$$\frac{\partial\left(uv - \frac{\tau_{xy}}{\rho}\right)}{\partial y} + \frac{\partial\left(uv - \frac{\tau_{xz}}{\rho}\right)}{\partial z} = gS \tag{16}$$

Where, $\tau_{xy} = \eta\frac{\partial u}{\partial y} - \rho\overline{u'v'}$, $\tau_{xz} = \eta\frac{\partial u}{\partial z} - \rho\overline{u'w'}$, and η is the kinetic viscosity.

In the two-dimensional flow, the first term on the left side of the equation in Eq. (16) should dominate and the second term has little effect and can be neglected. Eq. (16) can be simplified and integrated to obtain:

$$\mu\frac{du}{dy} - uv - \overline{u'v'} = -gSy + C \tag{17}$$

Substituting the boundary conditions, Eq. (17) can be obtained as follows:

$$\frac{du^+}{dy^+} - u^+v^+ - \frac{\overline{u'v'}}{u_*^2} = 1 - \frac{y}{H} \tag{18}$$

Where, $u^+ = u/u_*$, $y^+ = yu_*/\mu$, and $v^+ = v/u_*$.

A method for calculating the average turbulence structure is proposed here, where the probability of bursting phenomena r is first determined based on the direction of the vertical flow velocity. If the instantaneous velocity is greater than the average velocity, the subscript is recorded as u; otherwise, it is recorded as d. Where r is defined as:

$$r_u = \frac{\sum_0^{T_u}\Delta t_u}{T} \tag{19}$$

$$r_d = \frac{\sum_0^{T_d}\Delta t_d}{T} \tag{20}$$

Equation (19, 20) can be rewritten as:

$$r_u v_u + r_d v_d = 0 \tag{21}$$

The two-dimensional velocity affected by vegetation bursting can be defined as follows:

$$u_u = \frac{1}{T_u}\int_0^{T_u}\widetilde{u_u}dt \tag{22}$$

$$u_d = \frac{1}{T_d}\int_0^{T_d}\widetilde{u_d}dt \tag{23}$$

$$v_u = \frac{1}{T_u}\int_0^{T_u}\widetilde{v_u}dt \tag{24}$$

$$v_d = \frac{1}{T_d}\int_0^{T_d}\widetilde{v_d}dt \tag{25}$$

Where, $\widetilde{u}$ is the instantaneous velocity. In Eq. (18), the formula of Reynolds shear force can be obtained by introducing the bursting phenomenon.

$$\frac{du_u^+}{dy^+} - \frac{\overline{u_u'v_u'}}{u_{*u}^2} = \left(1 - \frac{y}{H}\right) + u_u^+v_u^+ \tag{26}$$

$$\frac{du_d^+}{dy^+} - \frac{\overline{u_d'v_d'}}{u_{*d}^2} = \left(1 - \frac{y}{H}\right) + u_d^+v_d^+ \tag{27}$$

Where, $u_u' = u - u_u$, $u_d' = \widetilde{u} - u_d$, $v_u' = \widetilde{v} - v_u$, and $v_d' = \widetilde{v} - v_d$ are the average turbulent structure; here $u_u^+ = u_u/u_{*u}$and $u_d^+ = u_d/u_{*d}$, u_{*u} and u_{*d} are the local frictional flow velocity in the case of upward turbulence and downward turbulence of the flow.

According to Yalin (1977), the viscosity of vortices in water can be expressed as follows:

$$\frac{v_T}{u_*} = \kappa\xi(1-\xi) \tag{28}$$

Where, v_t is the vortex viscosity, which can be approximated to be equal to μ, $\xi = y/H$, and κ is the Carmen constant. At $y = 0$, due to the non-slip boundary condition ($u_u = 0$ and $v_u = 0$), the influence of additional momentum flux u_uv_u on the velocity profile is negligible in the near-wall region. Substituting Eq. (28) into Eq. (26) and the non-slip boundary condition and simplify it:

$$\frac{du_u^+}{d\xi} = \frac{(1-\xi)}{\kappa\xi(1-\xi)} + \frac{u_u^+v_u^+}{\kappa\xi(1-\xi)} \tag{29}$$

Similarly, it can be obtained:

$$\frac{du_d^+}{d\xi} = \frac{(1-\xi)}{\kappa\xi(1-\xi)} + \frac{u_d^+v_d^+}{\kappa\xi(1-\xi)} \tag{30}$$

Equation (21) can be rewritten as:

$$v^+ = r_u v_u^+ u_{*u} + r_d v_d^+ u_{*d} = 0 \tag{31}$$

The vegetation influence is stochastic, so the average velocity gradient based on the probability of bursting phenomenon r can be expressed as:

$$u^+ = \frac{u}{u_*} = \frac{r_u u_u + r_d u_d}{u_*} = r_u u_u^+\frac{u_{*u}}{u_*} + r_d u_d^+\frac{u_{*d}}{u_*} \tag{32}$$

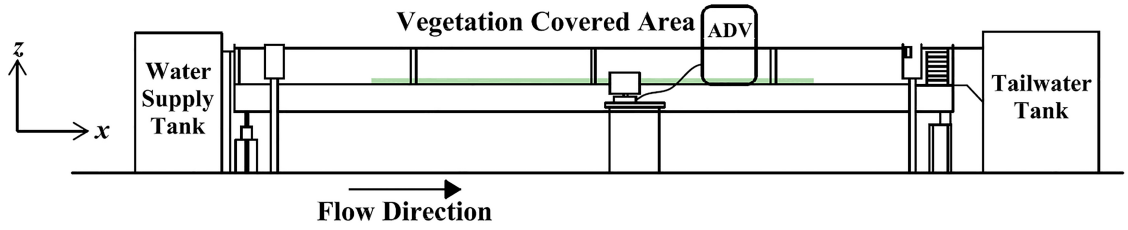


FIGURE 1 | Experimental platform for physical model of ecological channel.

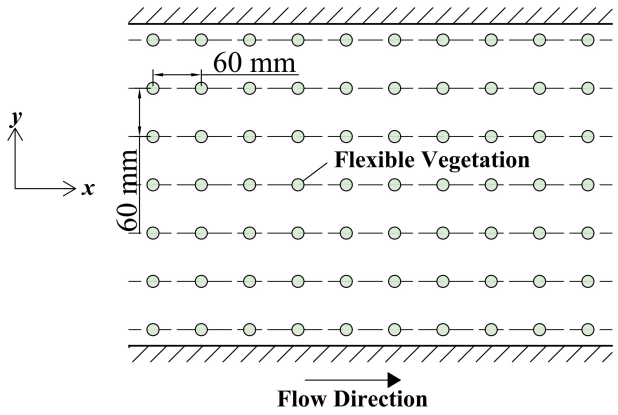


FIGURE 2 | Vegetation arrangement diagram.

So,

$$\frac{du^+}{d\xi} = r_u \frac{u_{*u}}{u_*} \frac{du_u^+}{d\xi} + r_d \frac{u_{*d}}{u_*} \frac{du_d^+}{d\xi} \tag{33}$$

By associating Eq. (33) with Eq. (29) and Eq. (30), the following results can be obtained:

$$\frac{du^+}{d\xi} = \frac{1}{\kappa\xi} + \frac{A}{\kappa\xi\,(1-\xi)} \tag{34}$$

Where,

$$A = r_u \frac{u_{*u}}{u_*} \left(u_u^+ v_u^+ - u_d^+ v_u^+\right) \tag{35}$$

Integrating Eq. (35) and simplifying it gives:

$$\frac{u}{u_*} = \frac{1}{\kappa} \ln \frac{y}{h_v} + \frac{A}{\kappa} \ln \frac{y}{h_v} - \frac{A}{\kappa} \ln \left(\frac{H-h_v}{H-y}\right) + \frac{u_{ud}}{u_*} \tag{36}$$

Equation (36) is the theoretical model of flow velocity distribution in open channels with flexible vegetation based on the bursting phenomenon. This velocity equation is named as the bursting phenomenon law (BPL).

EXPERIMENT

The model experiment was carried out on the open channel device in the hydraulic public test hall of the China Agricultural University. The rectangular channel is 6 m long, 0.8 m wide, 0.6 m high, and the slope is 5‰ (**Figure 1**). In order to ensure that the flow at the water inlet is as uniform as possible, the water inlet tank is aligned with the center line of the channel and a turbulent honeycomb is arranged at the water inlet of the tank. This experiment controls the water depth in the channel by controlling the valve system on the return water pipeline, so that the flow in the whole channel section can easily reach the uniform flow state in the open channel.

TABLE 1 | Experiment conditions.

Case	Stem spacing (m)	Vegetation density (flexible stems/m^2)	Flow Q (m^3/h)	Depth H (m)	Height of deflected stem h_v (m)
1	0.06	272	90	0.159	0.0551
2	0.06	272	100	0.170	0.0541
3	0.06	272	110	0.181	0.0533
4	0.06	272	120	0.191	0.0522

In this experiment, the flexible vegetation was simulated by plastic water plants; every plant was 0.06 m high, 0.03 m in diameter, and the total length of the vegetation section was 4 m. The vegetation is glued to the prefabricated perforated plastic plate, which is laid at the bottom of the sink. The vegetation strip was a single row of vegetation and equidistant (**Figure 2**); then, the position of the vegetation above the vegetation was measured with the plant as the reference.

The three-dimensional ultrasonic acoustic Doppler velocimetry (ADV) is used in this experiment to measure the velocity and ADV can measure the three-dimensional flow field information. The velocity range of ADV is from 0.001 to 4.5 m/s, the resolution can reach 0.0001 m/s, and the relative error is <1% of the measured velocity. To minimize the influence of ADV noise, the sample size of each measurement point in this experiment is about 2,000 times and the experiment time is 200 s. The downstream position is selected to ensure the stability of vegetation flow conditions.

The section of rectangular channel is symmetrical and the velocity distribution of the section is also symmetrical, so only half of the test data with the vertical line as the boundary line in the open channel need to be measured. Each measuring point is a predivided grid node. It is proposed to arrange one measuring line from the center vertical line to the right at every 0.04 m and

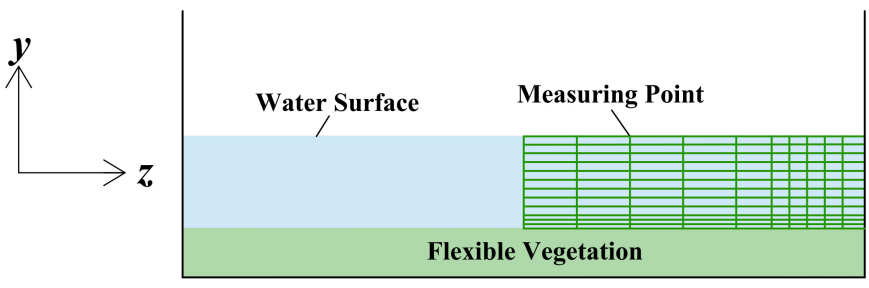


FIGURE 3 | Open channel mesh measuring.

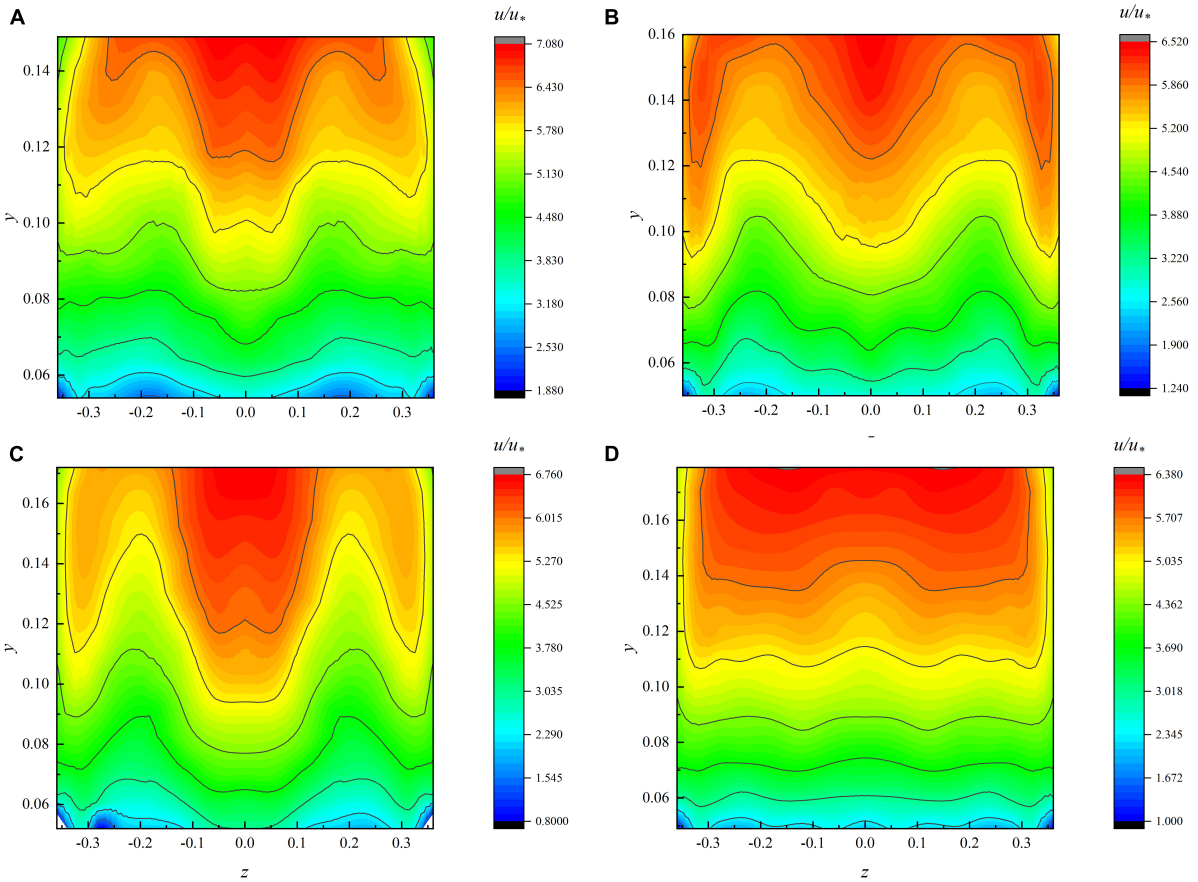


FIGURE 4 | Flow field contour. **(A)** 90 m^3/h, **(B)** 100 m^3/h, **(C)** 110 m^3/h, **(D)** 120 m^3/h.

one measuring point at every 0.01 m and to encrypt appropriately when close to vegetation (**Figure 3**), so as to ensure that the measured data can accurately reflect the actual flow field. Four flow conditions are set in the test. The specific test conditions are shown in **Table 1**.

RESULTS

The experiment is completed based on the experimental platform in section "Experiment" and the measured data under different flow conditions are drawn into the contour (**Figure 4**). It can be seen from **Figure 4** that the shape of the velocity curve is similar under each condition. The velocity distribution near the vertical line in the open channel generally decreases more evenly from the water surface to the vegetation as shown in **Figure 4A**; in other words, the maximum velocity point appears near the water surface. In the region near the sidewall, for the same z-coordinate of the velocimetric plumb line, it can be found that the velocity near the water surface is not the maximum and the maximum velocity point moves downward. There is also an appeal phenomenon in **Figures 4B–D**. It shows that it is not well considered to use a velocity formula to describe the velocity distribution of the whole section (Yang et al., 2004). Based on the antecedent analysis of contour, we use the experimental data to verify the theoretical formula derived in section "Modeling the Velocity Distribution in Channel With Flexible Vegetation."

Velocity Distribution in Channel Centerline

To explain the above phenomenon, the velocity distribution near the channel center line is discussed first. The velocity and depth are dimensionless, since they vary in size in each data series. By substituting the corresponding parameters into the VRL, the theoretical values of the flow velocity can be calculated. The theoretical values are plotted as straight lines, while the scatter points of the measured values are plotted in **Figure 5**. The scattered points are uniformly distributed around the straight line, indicating that the vegetated streambed flow distribution model is accurate in predicting the vertical distribution of flow near the center of the open channel containing flexible vegetation.

To prove the generalizability of the VRL, the velocity distribution measured by Yang and Choi (2009) is redrawn in **Figure 6**. Yang measured the velocity distribution in the center of the channel under the condition of uniform flow. Two different typical flow conditions (27 m^3/h and 37.8 m^3/h) are selected. The blue and red scattered points in **Figure 6** represent the

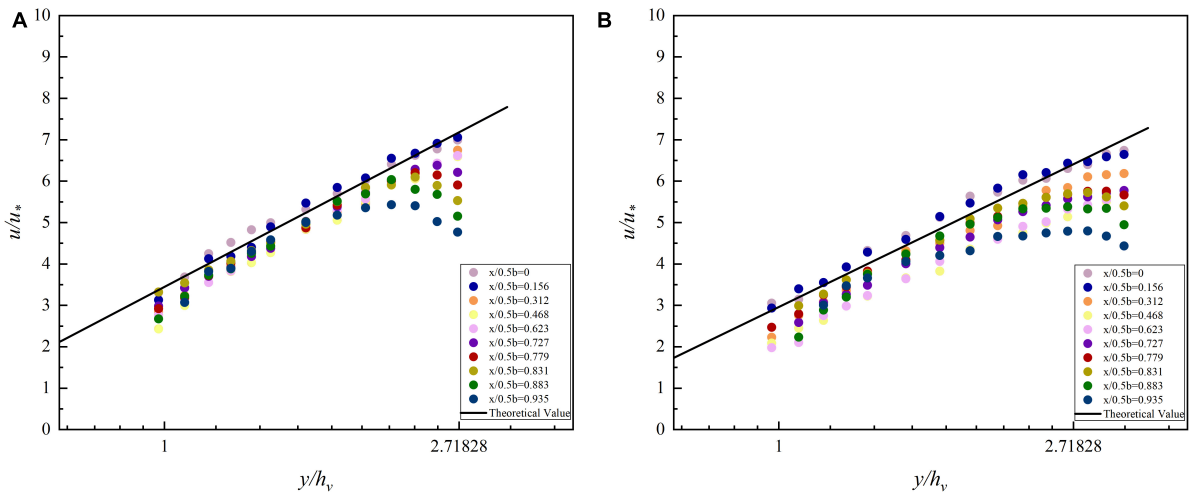


FIGURE 5 | Flow velocity distribution of different velocity measurement plumb lines in vegetated open channel cross sections. **(A)** 90 m^3/h, **(B)** 110 m^3/h.

dimensionless flow velocity under the two conditions and the red and blue lines represent the theoretical values calculated by applying the VRL under the corresponding conditions. **Figure 6** clearly shows that the theoretical values are in good agreement with the measured data, indicating that the VRL is generally applicable in the channel near the center of the channel is generally applicable.

However, in **Figure 5**, it is noticed that the velocity near the water surface deviates from the straight line in the near sidewall area. The velocity near the measured point changes from an upward trend to a downward trend, which is consistent with that shown in the flow field contour. It means that the VRL cannot be applied in this region. Next, the velocity distribution in this region is discussed by using the formula of bursting phenomenon.

Velocity Distribution of Channel Sidewall

This subsection focuses on the velocity distribution near the water surface.

In the experiment, it is found that the maximum point of velocity begins to move down from 0.1 m away from the sidewall and the VRL cannot well describe the velocity distribution in this area. It shows that the influence of vertical velocity should be considered. Therefore, the velocity distribution in this region should be calculated by the BPL.

The four-flow measurement vertical lines with the more obvious trend of decreasing flow velocity in the near sidewall area were selected for analysis under each operating condition. The theoretical value of velocity far from the water surface is calculated by the VRL and the BPL is used to calculate the area near the water surface. The theoretical values are plotted as black lines and the scattered points are the dimensionless velocities under each working condition, which are plotted in **Figure 7**. It is clear from **Figure 7** that the trend of the scatter is consistent with the curve. The average error between the theoretical and experimental values under different working conditions is below 4%. Take **Figure 7B** as an example, where the minimum error between the theoretical and experimental values is 0.1% and the maximum error is 9.2%. It shows that the model is more accurate to predict the velocity distribution near the water surface.

After the above analysis, it is shown that the flow velocity distribution of the whole section can be well described by using the VRL and the BPL in different areas.

DISCUSSION

In fact, it is because the flexible vegetation is inverted with the flow, which is equivalent to forming a "vegetation riverbed" with

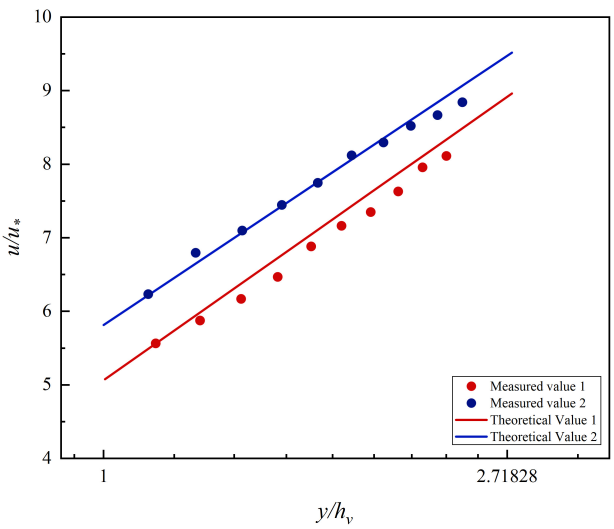


FIGURE 6 | Comparison of the velocity distribution of open channel with vegetation measured by Yang and the velocity distribution model of vegetation layer theory.

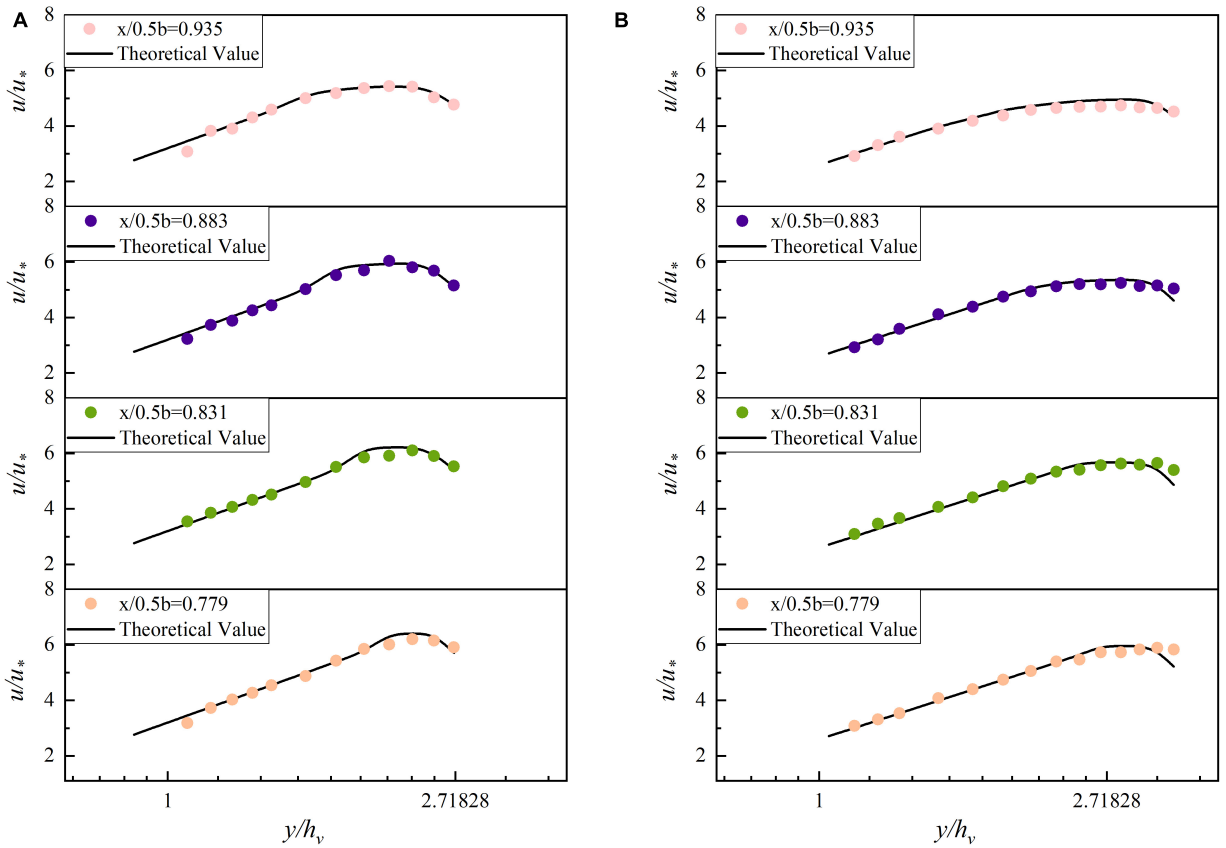


FIGURE 7 | Comparison of flow velocity analysis in the near sidewall area. **(A)** 90 m^3/h, **(B)** 120 m^3/h.

flexible vegetation as the bottom wall, so the flow in most non-vegetated areas still conforms to the logarithmic law. The biggest difference between the logarithmic velocity distribution formula of vegetated channel and smooth channel is the difference of "riverbed," which is also reflected in the VRL. In the VRL, it is obvious that the denominator of logarithmic term is the height of deflected vegetation (h_v), while the denominator of logarithmic term of traditional velocity distribution is the boundary layer separation point (Nikuradse, 1930). It shows that the height of deflected vegetation (h_v) is the characteristic length of open channel with vegetation. The characteristic length has also been mentioned in the study by Ghisalberti and Nepf (2002) and they believe that it can be expressed by using the momentum thickness. The momentum thickness needs to be obtained from the flow velocity calculations at the upper and lower boundaries of the mixed layer. In contrast, the calculation of the VRL is simpler.

Similarly, the VRL and the BPL have some similarities. The first term of the two equations is the same; it shows that the BPL also reflects the influence of flexible vegetation on flow. The difference appears in the following expressions. In the VRL, only one-dimensional mainstream velocity is used for calculation. The experimental results show that the flow near the center of the channel is closer to uniform flow and only the influence of the main flow direction can be considered. Most of the flow velocity equations proposed by researchers nowadays only consider the effect of flow velocity in the mainstream direction. For example, the flow velocity equation proposed by Nepf (2011) is partitioned for the vertical direction, but not for the cross-sectional flow velocity of the channel. However, there is a significant downward shift of the maximum velocity point in the region close to the sidewall. It is usually considered that the flow is influenced by the sidewall and produces a secondary flow. The manifestation of secondary flow is that there is a vortex near the water surface, which has a strong horizontal flow pointing to the center. This is the main reason for the location of the maximum flow velocity below the free surface (Nezu and Rodi, 1986). So, the vertical velocity should be included in the calculation. The BPL has a coefficient A, which includes the influence of mainstream velocity and vertical velocity.

Through the above analysis, it can be found that the flow velocity of open channel has obvious zoning phenomenon in the cross section. Take the maximum velocity point on the velocity measuring vertical line as the dividing point, next connect the dividing points on each vertical line and draw the red dividing line in **Figure 8**. The region above the dividing line is the region affected by the sidewall and the velocity in this region can be calculated by the BPL. The velocity in the area below the dividing line can be calculated by the VRL. The form of the theoretical partition line proposed by Daido (1992)

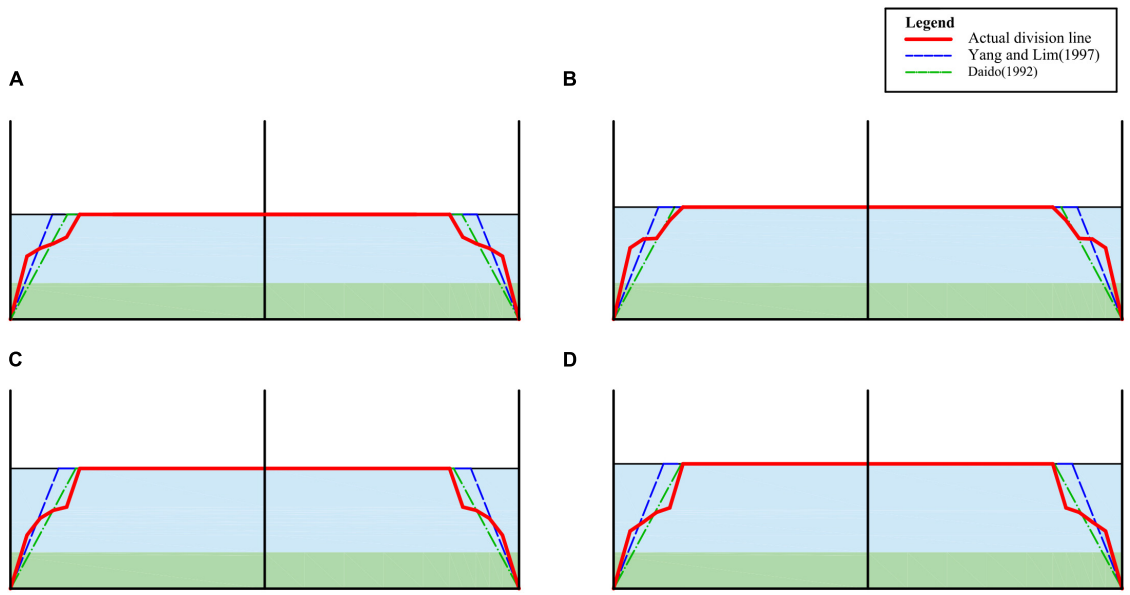


FIGURE 8 | Velocity dividing line of open channel with flexible vegetation. **(A)** 90 m^3/h, **(B)** 100 m^3/h, **(C)** 110 m^3/h, **(D)** 120 m^3/h.

and Yang and Lim (1997), who considered that the form of the dividing line is related to the roughness ratio of the sidewall to the sidewall and is plotted in their proposed theoretical dividing line in the figure (blue and brown lines). **Figure 8** shows that there is a gap between the theoretical dividing line and the actual dividing line clearly. The theoretical dividing line is a straight line near the sidewall, while the actual dividing line is a curve. Taking **Figure 8A** as an example, the dividing line in the near sidewall region has two particularly obvious inflection points, which is the same as that in **Figures 8B–D**. It shows that the split line is not a simple primary function form, but has a more complex function form, which will be discussed in depth in future studies.

CONCLUSION

This article studies the flow velocity distribution law with flexible vegetation theoretically and verifies the vegetation riverbed theory and the bursting phenomenon theory based on specific experimental data. The similarities and differences between the VRL and the BPL are also discussed. Based on the section dividing theory, the velocity dividing line of open channel with flexible vegetation is discussed. The main conclusions can be summarized as follows:

(1) Based on the vegetation riverbed theory, the velocity distribution formula of open channel with vegetation near the channel center can be derived. The equation shows that the vegetation in the open channel can be regarded as a "vegetation riverbed" and the flow in the upper layer of vegetation flows over the riverbed. The experimental data also confirm this hypothesis.

(2) The equation of flow velocity distribution in maximum velocity point drop was derived by using the bursting phenomenon theory. The BPL shows that the flow in the near sidewall area will be affected by secondary flow and the influence of the vertical flow velocity of the water needs to be considered. There are some similarities between the VRL and the BPL. Compared with the functional form of the VRL, the BPL has one more wake function. The correctness of the BPL is proved by the test data.

(3) There is an obvious zoning phenomenon in the flow velocity of the cross section of the open channel with flexible vegetation. The open channel velocity dividing partition line was found by the maximum velocity point of different velocity measurement plumb lines. Compared with the smooth open channel dividing line, the partition line with flexible vegetation has a more complex functional form. It provides a theoretical basis for the flow measurement of ecological open channel.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

SW contributed to the conceptualization, methodology, formal analysis, investigation, resources, data curation, writing—original draft, and visualization. YZ contributed to the investigation, resources, and data curation. TL contributed to the conceptualization, resources, writing—review and editing, and supervision. SL contributed to the investigation and resources. MZ contributed to the writing—review and editing. YH

contributed to the conceptualization, validation, writing—review and editing, supervision, and funding acquisition. 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. 51979275), Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University (Grant No. 19R06), Jilin Province Key Research and Development Plan Project (Grant No. 20180201036SF), Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University (Grant No. ICT20021), Chinese Universities Scientific Fund (Grant Nos. 2020TC033 and 10710301), and National Natural Science Foundation of China (Grant No. 51809106).

REFERENCES

Bennett, S. J., Pirim, T., and Barkdoll, B. D. (2002). Using simulated emergent vegetation to alter stream flow direction within a straight experimental channel. *Geomorphology* 44, 115–126. doi: 10.1016/S0169-555X(01)00148-9

Bouma, T. J., Duren, L. A. V., Temmerman, S., Claverie, T., Blanco-Garcia, A., Ysebaert, T., et al. (2005). Spatial flow and sedimentation patterns within patches of epibenthic structures: combining field, flume and modelling experiments. *Cont. Shelf Res.* 27, 1020–1045. doi: 10.1016/j.csr.2005.12.019

Daido, A. (1992). "Effect of Aspect ratio of channel and sediment density on transport phenomena," in *Proceedings of the International Symposium on Hydraulic Research in Nature and Laboratory*, Wuhan, B57–B72.

Ghisalberti, M., and Nepf, H. (2002). Mixing layers and coherent structures in vegetated aquatic flows. *J. Geophys. Res. Oceans* 107, 3.1–3.11. doi: 10.1029/2001JC000871

Helmiö, T. (2002). Unsteady 1D flow model of compound channel with vegetated floodplains. *J. Hydrol.* 269, 89–99.

Huai, W. X., Zeng, Y. H., Xu, Z. G., and Yang, Z. H. (2009). Three-layer model for vertical velocity distribution in open channel flow with submerged rigid vegetation. *Adv. Water Resour.* 32, 487–492. doi: 10.1016/j.advwatres.2008.11.014

Järvelä, J. (2002). Flow resistance of flexible and stiff vegetation: a flume study with natural plants. *J. Hydrol.* 269, 44–54. doi: 10.1016/S0022-1694(02)00193-2

Klopstra, D., Barneveld, H. J., Noortwijk, J., and Velzen, E. (1997). *Analytical Model for Hydraulic Roughness of Submerged Vegetation. Publication. 27th IAHR Congress*. HKV Consultants, San Francisco, 775–780.

Kubrak, E., Kubrak, J., and Rowinski, P. M. (2008). Vertical velocity distributions through and above submerged, flexible vegetation. *Hydrol. Sci. J.* 53, 905–920. doi: 10.1623/hysj.53.4.905

Leu, J. M., Chan, H. C., Jia, Y., He, Z., and Wang, S. S. Y. (2008). Cutting management of riparian vegetation by using hydrodynamic model simulations. *Adv. Water Resour.* 31, 1299–1308. doi: 10.1016/j.advwatres.2008.06.001

Lozanovska, I., Bejarano, M., Martins, M., Nilsson, C., Ferreira, M., and Aguiar, F. (2020). Functional diversity of riparian woody vegetation is less affected by river regulation in the Mediterranean than boreal Region. *Front. Plant Sci.* 11:857. doi: 10.3389/fpls.2020.00857

Nepf, H. (2011). Flow and transport in regions with aquatic vegetation. *Annu. Rev. Fluid Mech.* 44, 123–142. doi: 10.1146/annurev-fluid-120710-101048

Neumeier, U. (2005). Velocity and turbulence variations at the edge of saltmarshes. *Cont. Shelf Res.* 27, 1046–1059. doi: 10.1016/j.csr.2005.07.009

Nezu, I., Nakagawa, H., and Jirka, G. H. (1993). Turbulence in open-channel flows. *J. Hydraulic Eng.* 120, 1235–1237.

Nezu, I., and Rodi, W. (1986). Open-channel flow measurements with a laser doppler anemometer. *J. Hydraulic Eng.* 112, 335–355.

Nikuradse, J. (1930). Untersuchungen über turbulente Strömungen in nicht kreisförmigen Rohren. *Ing. Arch.* 1, 306–332. doi: 10.1007/BF02079937

Okamoto, T., Nezu, I., and Sanjou, M. (2016). Flow-vegetation interactions: length-scale of the "monami" phenomenon. *J. Hydraulic Res.* 54, 251–262. doi: 10.1080/00221686.2016.1146803

Prandtl, L. (1925). Uber die Ausgebildete Turbulenz. *Z. Angew. Math. Mech.* 1, 431–436.

Pu, J. H., Hussain, A., Guo, Y. K., Nikolaos, V., Prashanth, R., Hanmaiahgari, B., et al. (2019). Submerged flexible vegetation impact on open channel flow velocity distribution: an analytical modelling study on drag and friction. *Water Sci. Eng.* 012, 121–128. doi: 10.1016/j.wse.2019.06.003

Righetti, M., and Armanini, A. (2002). Flow resistance in open channel flows with sparsely distributed bushes. *J. Hydrol.* 269, 55–64. doi: 10.1016/s0022-1694(02)00194-4

Rivaes, R., Pinheiro, A., Egger, G., and Ferreira, M. (2017). The Role of river Morphodynamic disturbance and groundwater hydrology as driving factors of riparian landscape Patterns in Mediterranean Rivers. *Front. Plant Sci.* 8:1612. doi: 10.3389/fpls.2017.01612

Siniscalchi, F., Nikora, V. I., and Aberle, J. (2012). Plant patch hydrodynamics in streams: mean flow, turbulence, and drag forces. *Water Resour. Res.* 48:W01513. doi: 10.1029/2011WR011050

Tracy, H. J., and Lester, C. M. (1961). *Resistance Coefficients and Velocity Distribution, Smooth Rectangular Channel. U.S. Geological Survey (USGS) Water-Supply Paper 1592-A*. (Washington, DC: USGS), A1–A18.

Wang, C., Fan, X. L., Wang, P. F., Hou, J., and Qian, J. (2016). Flow characteristics of the wind-driven current with submerged and emergent flexible vegetations in shallow lakes. *J. Hydrodyn. Ser. B* 28, 746–756. doi: 10.1016/S1001-6058(16)60677-7

Wang, W. J., Huai, W. X., Zeng, Y. H., and Zhou, J. F. (2015). Analytical solution of velocity distribution for flow through submerged large deflection flexible vegetation. *Appl. Math. Mech.* 36, 107–120. doi: 10.1007/s10483-015-1897-9

Wilson, C. A. M. E., Yagci, O., Rauch, H. P., and Olsen, N. R. B. (2005). 3D numerical modelling of a willow vegetated river/floodplain system. *J. Hydrol.* 327, 13–21. doi: 10.1016/j.jhydrol.2005.11.027

Yalin, M. S. (1977). *Mechanics of Sediment Transport*. Oxford: Pergamon Press.

Yang, S. Q. (2010). Conditionally averaged turbulent structures in 2D channel flow. *Water Manage.* 163, 79–88. doi: 10.1680/wama.2010.163.2.79

Yang, S. Q., and Lim, S. Y. (1997). Mechanism of energy transportation and turbulent flow in a 3D channel. *J. Hydraulic Eng.* 123, 684–692.

Yang, S. Q., Tan, S. K., and Lim, S. Y. (2004). Velocity distribution and dip-phenomenon in smooth uniform open channel flows. *J. Hydraulic Eng.* 130, 1179–1186.

Yang, W., and Choi, S. U. (2009). Impact of stem flexibility on mean flow and turbulence structure in depth-limited open channel flows with submerged vegetation. *J. Hydraulic Res.* 445–454. doi: 10.1080/00221686.2009.9522020

Zong, L., and Nepf, H. (2009). Flow and deposition in and around a finite patch of vegetation. *Geomorphology* 116, 363–372. doi: 10.1016/j.geomorph.2009.11.020

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fpls.2021.770217

PocketMaize: An Android-Smartphone Application for Maize Plant Phenotyping

Lingbo Liu[1], Lejun Yu[1,2], Dan Wu[1], Junli Ye[3], Hui Feng[3,4], Qian Liu[1,2] and Wanneng Yang[3,4]**

[1] Wuhan National Laboratory for Optoelectronics, Britton Chance Center for Biomedical Photonics, Key Laboratory of Ministry of Education for Biomedical Photonics, Department of Biomedical Engineering, Huazhong University of Science and Technology, Wuhan, China, [2] School of Biomedical Engineering, Hainan University, Haikou, China, [3] National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan, China, [4] Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China*

OPEN ACCESS

Edited by:
Dongjian He, Northwest A&F University, China

Reviewed by:
Honghua Jiang, Shandong Agricultural University, China
Urs Schmidhalter, Technical University of Munich, Germany

***Correspondence:**
Qian Liu
qliu@hainanu.edu.cn
Wanneng Yang
ywn@mail.hzau.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *03 September 2021*
Accepted: *05 November 2021*
Published: *25 November 2021*

Citation:
Liu L, Yu L, Wu D, Ye J, Feng H, Liu Q and Yang W (2021) PocketMaize: An Android-Smartphone Application for Maize Plant Phenotyping. Front. Plant Sci. 12:770217. doi: 10.3389/fpls.2021.770217

A low-cost portable wild phenotyping system is useful for breeders to obtain detailed phenotypic characterization to identify promising wild species. However, compared with the larger, faster, and more advanced in-laboratory phenotyping systems developed in recent years, the progress for smaller phenotyping systems, which provide fast deployment and potential for wide usage in rural and wild areas, is quite limited. In this study, we developed a portable whole-plant on-device phenotyping smartphone application running on Android that can measure up to 45 traits, including 15 plant traits, 25 leaf traits and 5 stem traits, based on images. To avoid the influence of outdoor environments, we trained a DeepLabV3+ model for segmentation. In addition, an angle calibration algorithm was also designed to reduce the error introduced by the different imaging angles. The average execution time for the analysis of a 20-million-pixel image is within 2,500 ms. The application is a portable on-device fast phenotyping platform providing methods for real-time trait measurement, which will facilitate maize phenotyping in field and benefit crop breeding in future.

Keywords: smartphone, application, plant phenotyping, deep learning, maize plants

INTRODUCTION

Maize (Zea mays L) is one of the essential crops cultivated primarily for food, animal feed, and biofuel, and a more significant amount of maize by weight is produced each year than any other grain (Ritchie and Roser, 2020). Maize plant traits, such as plant architecture, plant biomass, plant projected area, and plant height, are essential factors in the study of maize biology, growth analysis, and yield estimation (Golzarian et al., 2011). Leaves are the primary photosynthetic organs and fundamental importance to maize, acting as transporters, carrying essential materials and energy from the environment, and eliminating waste products (Efroni et al., 2010). Thus, leaf traits, such as leaf area, leaf shape, and leaf number, are also of great significance in maize breeding (Yang et al., 2013). Moreover, the traits of each individual leaf at different heights contribute differently to the final yield prediction (Zhang et al., 2017).

Wild species related to agricultural crops (CWR, crop wild relatives) represent a large pool of genetic diversity, providing new allelic variation for yield improvements, disease resistance, farming practices, and market demands (Dempewolf et al., 2017). The advent of next-generation sequencing technology has resulted in a significant improvement in genomics (Koboldt et al., 2013) and implemented high-throughput genome sequencing for CWR. However, there are substantial gaps in accessible CWR in gene banks, and available evidence indicates that the crop diversity present in farmers' fields has declined, leading to the rareness or even disappearance of many farmers' varieties and landraces (FAO, 2010; Pilling et al., 2020). Among the most critical crops across the global food supply, such as wheat, rice, and soybean, maize wild relatives gain the highest priority for further collection to improve their representation in gene banks (Castañeda-Álvarez et al., 2016). When breeders collect CWR resources, a portable device that can provide detailed phenotypic characterization on device in wild conditions is urgently needed.

Over the past few decades, many versatile and high-throughput phenotyping platforms have been developed (Yang et al., 2020). Compared with other phenotyping trait collection methods, image-based phenotyping is noninvasive, scalable, and easy to automate (Das Choudhury et al., 2016). Granier et al. (2006) developed one of the first automated visible-light imaging systems called PHENOPSIS for detecting Arabidopsis responses to water deficit in 2003. Walter applied soil-filled rhizoboxes to make the root visible and established GROWSCREEN for both aboveground and belowground phenotyping in 2007 (Walter et al., 2007). Later, a chlorophyll fluorescence imaging system was attached to the platform, and GROWSCREEN was updated into GROWSCREEN FLUORO, allowing the phenotyping of leaf growth and chlorophyll (Jansen et al., 2009). In the next few years, larger-scale phenotyping platforms in the laboratory, such as Phenoscope (Tisné et al., 2013) and Phenovator (Awlia et al., 2016), were designed for potted plants. These platforms combined the rotating imaging table for multiangle imaging, a high-speed x–y rail system for camera movement (camera to plant) or plant movement (plant to camera), and a dark acclimation chamber for a more stable imaging environment. In general, phenotyping systems in the laboratory are rapidly developing and contain more advanced sensors for additional traits unable to be acquired before. However, indoor phenotyping platforms are costly, time-consuming, immovable, and require skilled engineers for maintenance. To provide phenotyping measurements in the field, a portable, simple-to-operate, and cost-effective phenotyping platform is needed.

Taking advantage of advances in sensors and chip computation power, modern smartphones have become a new solution that combines sensors, platforms, and processing, and a few methodologies for phenotyping with smartphones have been developed (Araus and Kefauver, 2018). The fractional vegetation cover can be estimated from simple calculations with traditional RGB images taken above crop canopies using the smartphone's own processing capacities (Patrignani and Ochsner, 2015; Chung et al., 2017). PocketPlant3D uses the device accelerator and magnetometer to measure the leaf insertion angle and the leaf angles from the insertion to the tip (Confalonieri et al., 2017). PocketLAI acquires real-time images from below the plant canopy. It uses the smartphone accelerator to obtain the smartphone's current depression angle and detect sky pixels when the angle between the vertical and the normal to the screen reaches 57° to estimate plant LAI (Orlando et al., 2016). PocketN estimates plant nitrogen content from digital images (Confalonieri et al., 2015). The iPad application "Estimate" takes images of a single expanded leaf and uses standard area diagrams (SADs) to estimate the severity of Cercospora leaf spot (Pethybridge and Nelson, 2018). Some researchers develop applications to acquire images and send them to a server for advanced data processing to transfer machine learning approaches to smartphone applications (Singh et al., 2018). This client-server architecture fills the smartphone computation capacity gap by transmitting image data to an in-house server for advanced image processing to detect Cercospora leaf spots on sugar beet (Hallau et al., 2018). A cloud-based system that can send the images taken from the greenhouse to the cloud is developed for water stress prediction using window-based support vector regression (multimodal SW-SVR) (Kaneda et al., 2017). These applications revealed the potential of mobile devices in plant phenotyping.

In previous work, our phenotyping team developed a high-throughput indoor phenotyping facility called HRPF to extract rice phenotypic traits (Yang et al., 2014), and more novel imaging techniques were renewed and applied in more crops, such as multiangle RGB imaging for 3D reconstruction of wheat plants (Fang et al., 2016), hyperspectral imaging for rice plant biomass (Feng et al., 2013) and rice leaf chlorophyll (Feng et al., 2013), and micro-CT for rice tiller traits (Wu et al., 2019). In the present work, we developed a portable on-device phenotyping system running on Android to nondestructively extract 15 plant traits, 25 leaf traits and 5 stem traits with high efficiency (up to 3 s per plant), which provides a real-time quantitative maize trait analysis for breeders.

MATERIALS AND METHODS

Material and Experimental Design

The study area was located at Huazhong Agricultural University, Wuhan, Hubei Province, China (30.5N, 114.3E) at an average elevation of 16 m. The maize variety of JinZhongYu (YT0213/YT0235) was sown and germinated during the summer of 2015. Ninety maize plants were planted in a plastic pot and in the field. The pots were 23.5 cm in diameter and 35 cm in height with approximately 6 l of experimental soil (pH 5.45, total nitrogen 0.241 g kg^{-1}, total potassium 7.20 g kg^{-1}, total phosphorus 0.74 g kg^{-1}, alkali-hydrolysable nitrogen 144.06 mg kg^{-1}, available potassium 188.64 mg kg^{-1}, available phosphorus 16.81 mg kg^{-1}, organic matter 46.55 g kg^{-1}).

The measurement started 30 days after sprouting. Every 3 days, nine plants in the pots and nine plants in the field were randomly picked and photographed outdoors via an application (PocketMaize) running on an ANDROID smartphone (MEIZU, MX4). A portable black backdrop is placed behind the plant as background, and a marker object is placed next to the plant to

calculate the resolution between image pixels and the real world. Four images with pitch angles of approximately 0° (front view), 10°, 15°, and 20° were taken for each plant with arbitrary distance from camera to the plant and imaging height. Images were stored in JPG format with a resolution of 3,936 × 5,248 pixels. The app stores the pictures and records the current spatial angle, time, and date when a picture is taken. Other necessary pieces of information, such as plant ID, can also be manually input.

After imaging, plant height was manually measured with a ruler vertically placed against the plastic pot's edge or the field ground on top of the soil surface. The shoot part of the maize plant was then cropped down for destructive measurements. The fresh leaf biomass and stem biomass were estimated separately. The individual leaves of each plant were cut, and the leaf area was measured using high-throughput leaf scoring (HLS) (Yang et al., 2015). Then, the plants were sealed and oven-dried for further dry-weight determination.

The Image Process Procedure

In this study, we developed an application running on Android smartphones (called PocketMaize) for image acquisition, image processing, and plant traits extraction combined (**Figure 1**). Image processing's key steps include image calibration, angle calibration, image segmentation, skeletonization, stem and leaf extraction, and phenotypic traits calculation.

The first step, camera distortion calibration using an OpenCV calibration function (Zhang, 2000), is optional. A black and white calibration pattern pasted on a plastic plate was used to obtain 20–25 images. Furthermore, the imaging angles between each image should have apparent differences to ensure accuracy.

The second step is to calibrate the rotation angle and depression angle. As shown in **Figures 2A,B**, the depression angle α is the angle between the normal n to the plane on which the device's screen lays and the horizontal plane, while the rotation angle γ is the angle between the y-axis of the screen and the zenith.

An ideal image for trait extraction should be perpendicular to the ground and have the same object-pixel resolution for the whole image. One of the best options is orthographic projection imaging. For most of the other image-based phenotyping systems, camera lens distortion calibration is sufficient because in these systems, the cameras are fixed to obtain a stable imaging angle. However, in our application, the position and direction of the camera are continuously changing, making it essential to calibrate ordinary images with different rotation and depression angles to an approximate orthographic projection image by image transformation.

Gravity sensors in the smartphone provide live data of the rotation angle and depression angle, and these two angles are calibrated separately. **Figure 1A** shows an original image obtained from camera calibration whose rotation angle and depression angle need to be calibrated. Usually, the perpendicularity is satisfied by simply rotating the image clockwise or anticlockwise using the rotation angle obtained from the accelerometer to match the gravity direction while a perspective transform was applied to adjust the depression angle. **Figures 1B,C** display the results of the rotation angle calibration and depression angle calibration, respectively.

The perspective transform is used in depression angle calibration, which is a nonlinear geometric transformation that can change an image from one viewpoint to another viewpoint or, in other words, change the position of the image plane. It is widely used in image processing, including length calculation, marking recognition in images (Liu et al., 2012) and vision guidance for vehicles (Torii, 2000). Perspective transform can transform the ordinary images taken in this experiment with different imaging angles to an approximate orthographic projection image. It needs the coordinates of four sets of points, four given points on the original image plane and their corresponding points on the target image plane, to calculate the perspective transform matrix. Placing four markers on to the background can reduce obstacles in image processing. The markers might be obscured by leaves, and it is troublesome to determine the relative positions in the wild. Therefore, a camera calibration method without markers was developed.

Examine a 2W × 2H rectangle with four symmetric corner points LT, RT, LB, RB on the target image plane (object plane) and their corresponding points $LT^{'}$, $RT^{'}$, $LB^{'}$, $RB^{'}$ on the origin image plane. **Figure 2C** shows the position of these eight points, and point C is the origin point of the coordinate system on both the object plane and image plane. The coordinates of these eight points are $LT(-W, H)$, $RT(W, H)$, $LB(-W, -H)$, and $RB(W, H)$ on the target image plane and $LT^{'}\left(-W^{'}_{upper}, H^{'}_{upper}\right)$, $RT^{'}\left(W^{'}_{upper}, H^{'}_{upper}\right)$, $LB^{'}\left(-W^{'}_{lower}, H^{'}_{lower}\right)$, and $RB^{'}\left(W^{'}_{lower}, H^{'}_{lower}\right)$ on the origin image plane. **Figure 2D** is the longitudinal section at the center, while T and $T^{'}$ are the center of LT, RT and $LT^{'}$, $RT^{'}$ and B and $B^{'}$ are the center of LB, RB and $LB^{'}$, $RB^{'}$. First, for the upper part of the image, let's mark

$$\begin{cases} L_{upper} &= \overline{OT} \\ L^{'}_{upper} &= \overline{OT^{'}} \end{cases} \tag{1}$$

Then, we have

$$\frac{W^{'}_{upper}}{W} = \frac{L^{'}_{upper}}{L_{upper}} \tag{2}$$

Let β be the actual viewing angle of the point and D be the distance between the camera and the plant; then, we have:

$$H^{'}_{upper} = D\tan\beta \tag{3}$$

$$H^{'}_{max} = D\tan\beta_{max} \tag{4}$$

$$\beta = \tan^{-1}\left(\frac{H^{'}\tan\beta_{max}}{H^{'}_{max}}\right) \tag{5}$$

where β_{max} is the half vertical field of view (VFOV) of the camera and $H^{'}_{max}$ is the half y resolution of the camera.

The trigonometric relationship in the upper half of the image can be described as follows:

$$D = L^{'}_{upper}\cos\beta \tag{6}$$

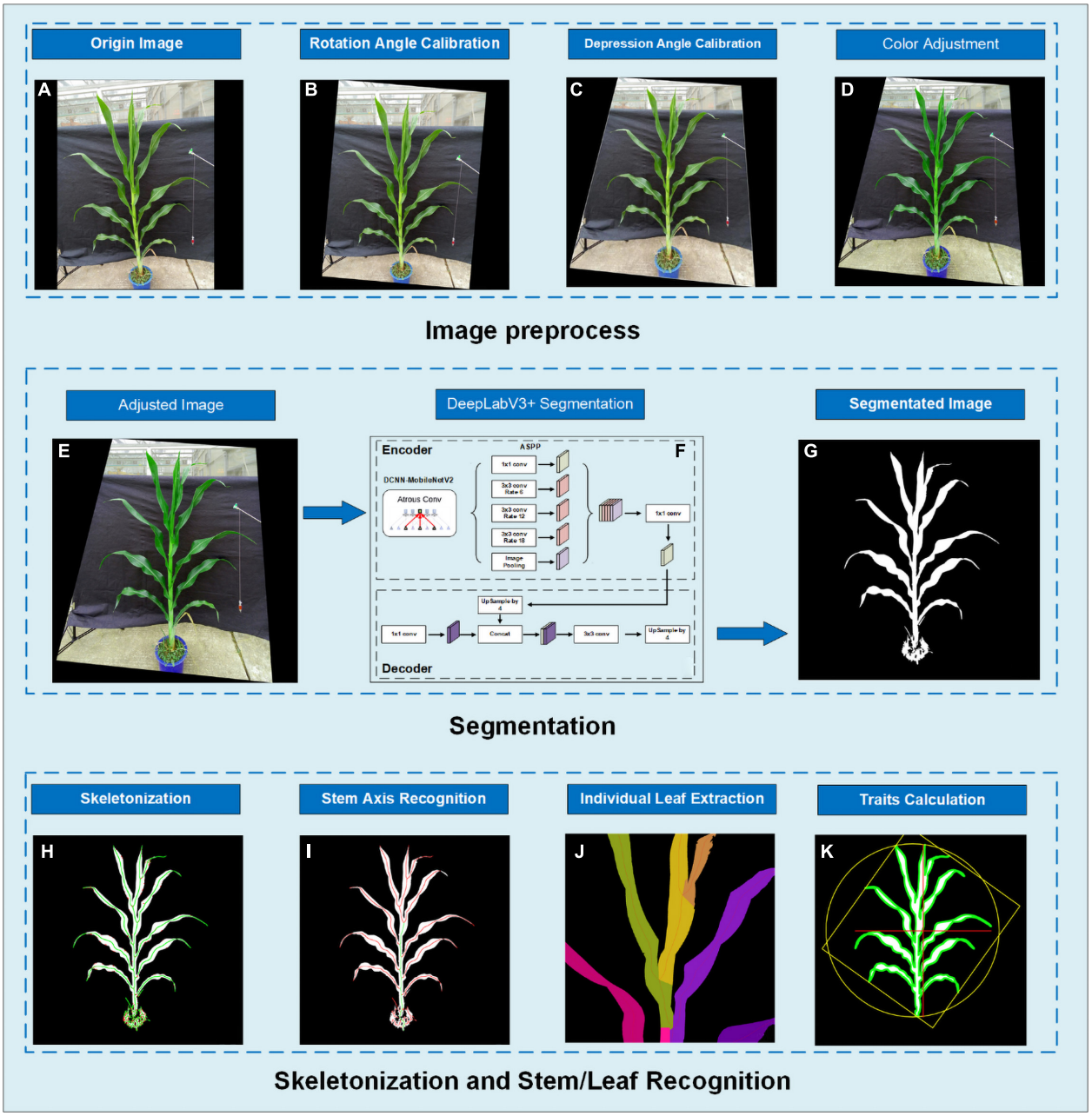


FIGURE 1 | The image analysis pipeline showing **(A)** original image; **(B)** rotated image to calibrate rotation angle; **(C)** depression angle calibration using perspective transform; **(D)** color adjustment; **(E)** the resulting image of preprocessing; **(F)** segmentation using DeepLabV3+; **(G)** segmentation result image; **(H)** skeletonization using our distance transform-based algorithm; **(I)** stem axis recognition by finding the overlaid route; **(J)** pixel extraction for each individual leaf; **(K)** result image.

$$\frac{H_{upper}}{\sin\beta} = \frac{D}{\sin\left[\pi-\beta-\left(\frac{\pi}{2}-\alpha\right)\right]} \tag{7}$$

$$\frac{L_{upper}}{\sin\left(\frac{\pi}{2}-\alpha\right)} = \frac{D}{\sin\left[\pi-\beta-\left(\frac{\pi}{2}-\alpha\right)\right]} \tag{8}$$

At last, we have

$$L'_{upper} = \frac{\cos(\alpha-\beta)}{\cos\alpha\cos\beta} L_{upper} \tag{9}$$

and

$$H'_{upper} = \frac{\cos(\alpha-\beta)}{\cos\beta} H_{upper} \tag{10}$$

Similarly, for the lower half of the image, we have

$$\frac{H_{lower}}{\sin\left[\pi-\alpha-\left(\frac{\pi}{2}-\alpha-\beta\right)\right]} = \frac{H'_{lower}}{\sin\left(\frac{\pi}{2}-\alpha-\beta\right)} \tag{11}$$

$$L'_{lower} = \frac{H'_{lower}}{\sin\beta} \tag{12}$$

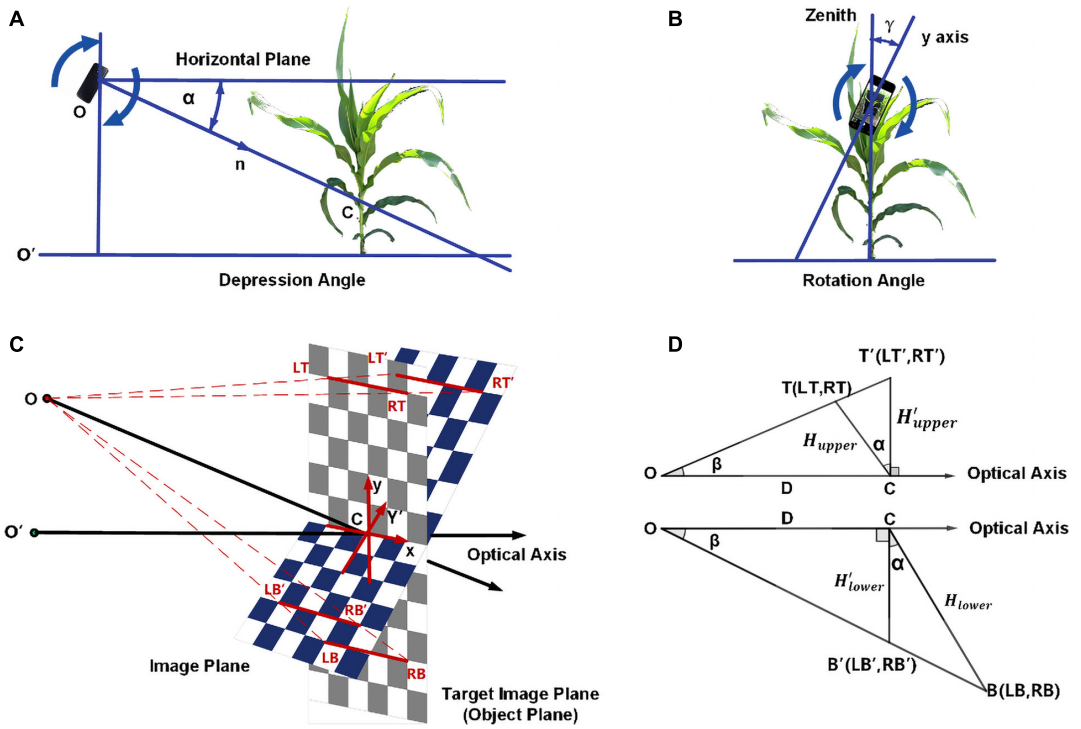


FIGURE 2 | The calibration of rotation angle and depression angle showing **(A)** explanation of depression angle; **(B)** explanation of rotation angle; **(C)** schematic diagram for depression angle; **(D)** longitudinal section for depression angle.

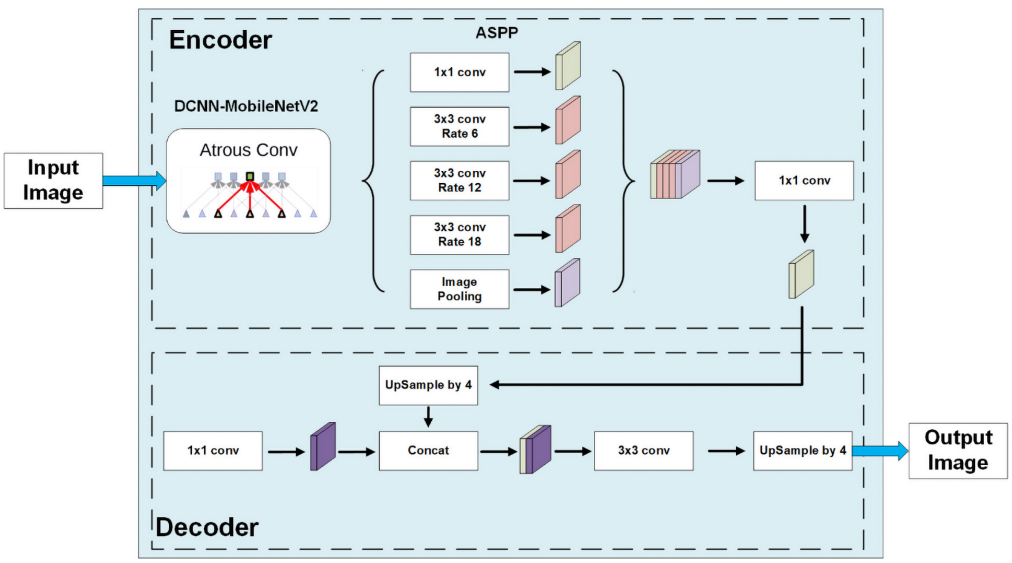


FIGURE 3 | The model structure of DeepLabV3+ with the MobuleNetV2 backbone.

$$\frac{L_{lower}}{\sin\left(\frac{\pi}{2}+\alpha\right)} = \frac{H_{lower}}{\sin\beta} \tag{13}$$

Hence, the final proportion is given by

$$H'_{lower} = \frac{\cos(\alpha+\beta)}{\cos\beta} H_{lower} \tag{14}$$

$$L'_{lower} = \frac{\cos(\alpha+\beta)}{\cos\alpha\cos\beta} L_{lower} \tag{15}$$

From formulas (9), (10), (14), and (15), for any given point on the original image plane, we can calculate the coordinates of the corresponding points on the target image plane and

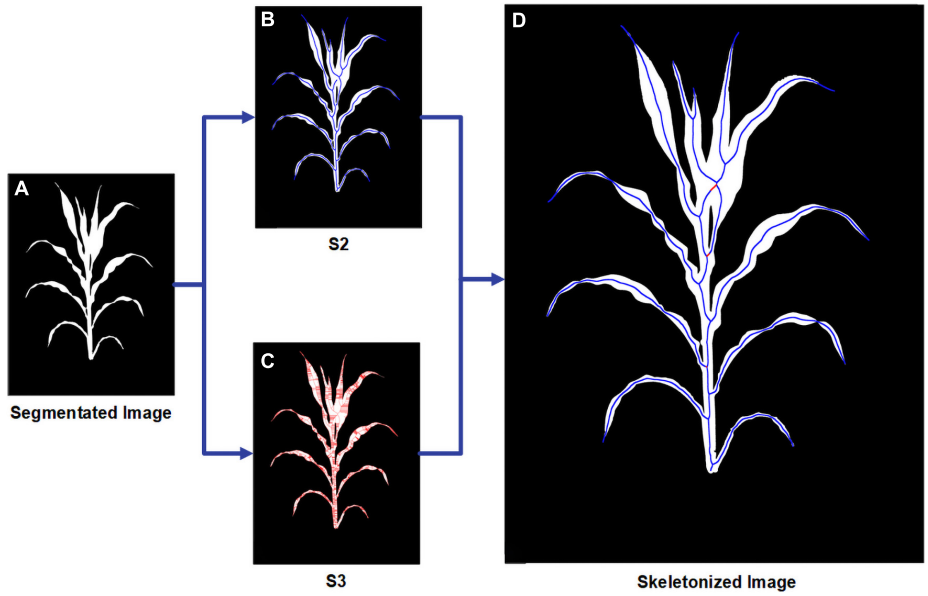


FIGURE 4 | Schematic workflow for the skeletonization procedure showing **(A)** origin segmented image; **(B)** all points in S_2; **(C)** all points in S_3; **(D)** final result for skeletonization with the points from S_2 marked in blue and the points from S_3 marked in red.

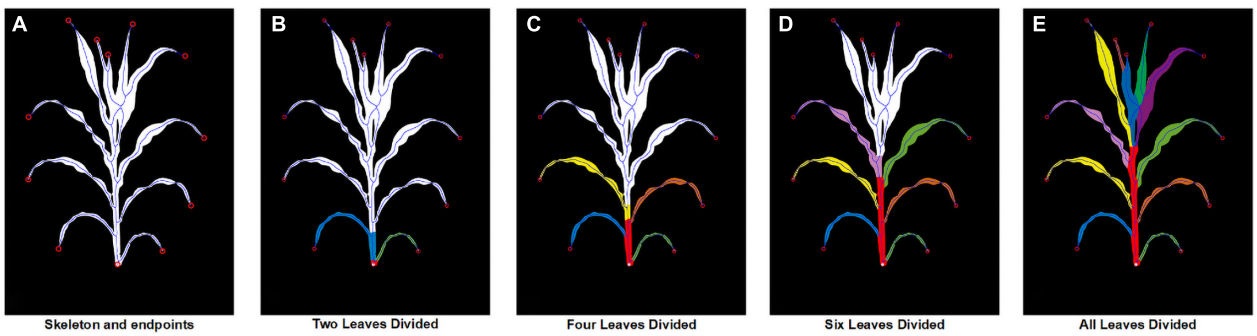


FIGURE 5 | The procedure of leaf pixel segmentation showing **(A)** the original segmented image and skeletonized image; **(B–E)** the segmentation result of two, four, six and all leaves with each leaf marked in different colors and the stem marked in red.

vice versa. Then, we can calculate the eight points needed to transform the ordinary image to an approximate orthographic projection image.

The color enhancement in this manuscript aims to standardize the image color according to the main color of the target object and enlarge the color difference between the plant part and background. The purpose of color enhancement here is to improve the segmentation result of DeepLabV3+ under different surrounding weather and illumination conditions.

Use DeeplabV3+ With the MobileNet Backbone for Segmentation

Segmentation of the plant image is the critical step for the image process. We introduced the DeepLabV3+ model with a MobileNet backbone to obtain segmented results of images with different lightness conditions and backgrounds. DeepLabV3+ is a convolutional neural network model designed for pixel-based semantic image segmentation that has three improved versions (Chen et al., 2018).

Convolutional neural networks use several layers of filters convolved with the input data to greatly reduce the dimension of input data and extract features of the image. These features from each layer are combined into feature maps that can then be used to make the output prediction. Compared with other convolutional neural networks, DeepLabV1 (Chen et al., 2016) introduced a dilated convolution to increase the receptive field to regain the data lost in the pooling layer and used the conditional random field (CRF) to improve boundary recognition. DeepLabV2 (Chen et al., 2017) established the model with atrous spatial pyramid pooling (ASPP) to handle images of similar objects with different scales. DeepLabV3 (Chen et al., 2018) adds a batch normalization layer into the ASPP, and DeepLabV3+ uses a simple decoder module to further upgrade

TABLE 1 | Trait classification and abbreviation.

Trait classification	Trait	Trait abbreviation
Plant traits	Maximum plant height in side view	MPH
	Vertical plant height in side view	VPH
	Plant width in side view	PW
	Total projected area	TPA
	Green projected area/total projected area in side view	GPAR
	Total projected area/bounding rectangle area ratio in side view	TBR
	Plant perimeter in side view	PP
	Perimeter/projected area ratio in side view	PAR
	Plant compactness in side view	PC
	Fractal dimension in side view	FD
	Height to width ratio of minimum circumscribed box in side view	HWR
	The area of convex hull	ACH
	The perimeter of convex hull	PCH
	Plant area/convex hull area	PCHAR
	Total dry weight	TDW
Leaf traits	Total leaf dry weight	LDW
	Total leaf area	TLA
	Total leaf projection area	TLPA
	Total leaf length per plant	TLL
	Leaf number per plant	LN
	Standard deviation of straightened leaf length per plant	SDSLL
	Average distance between the leaf tip and node per plant	LNL
	Standard deviation of the distance between the leaf tip and node per plant	SDLNL
	Average leaf curvature per plant	LC
	Standard deviation of leaf curvature per plant	SDLC
	Average leaf tangency angle per plant	LTA
	Standard deviation of leaf tangency angle per plant	SDLTA
	Average leaf straight angle per plant	LSA
	Standard deviation of leaf straight angle per plant	SDLSA
	Average straightened leaf length in lower half of plant	SLL_below
	Average distance between the leaf tip and node in lower half of plant	LNL_below
	Average leaf curvature in lower half of plant	LC_below
	Average of leaf tangency angle in lower half of plant	LTA_below
	Average of leaf straight angle in lower half of plant	LSA_below
	Average straightened leaf length in upper half of plant	SLL_above
	Average distance between the leaf tip and node in upper half of plant	LNL_above
	Average leaf curvature in upper half of plant	LC_above
	Average of leaf tangency angle in upper half of plant	LTA_above
	Average of leaf straight angle in upper half of plant	LSA_above
	Total leaf dry weight	TLDW
Stem traits	Stem height	SH
	Stem projection area	SPA
	Average stem width	SW
	Stem volume	SV
	Stem dry weight	SDW

boundary recognition. The DeepLabV3 Plus model is a deep convolutional neural network with atrous convolution that can increase the receptive field without increasing the number of parameters or reducing the dimension of space.

Several kinds of backbones can be used in DeepLab, including ResNet (He et al., 2015), Xception (Chollet, 2017) and MobileNet (Howard et al., 2017). All these models have good performance in maize segmentation, especially Xception. However, to transfer the model to mobile devices, we decided to train our DeepLab model with the most lightweighted MobileNet. The structure of the whole DeepLabV3+ model with the MobileNetV2 backbone is shown in **Figure 3**. These modules are implemented in TensorFlow (Abadi et al., 2016).

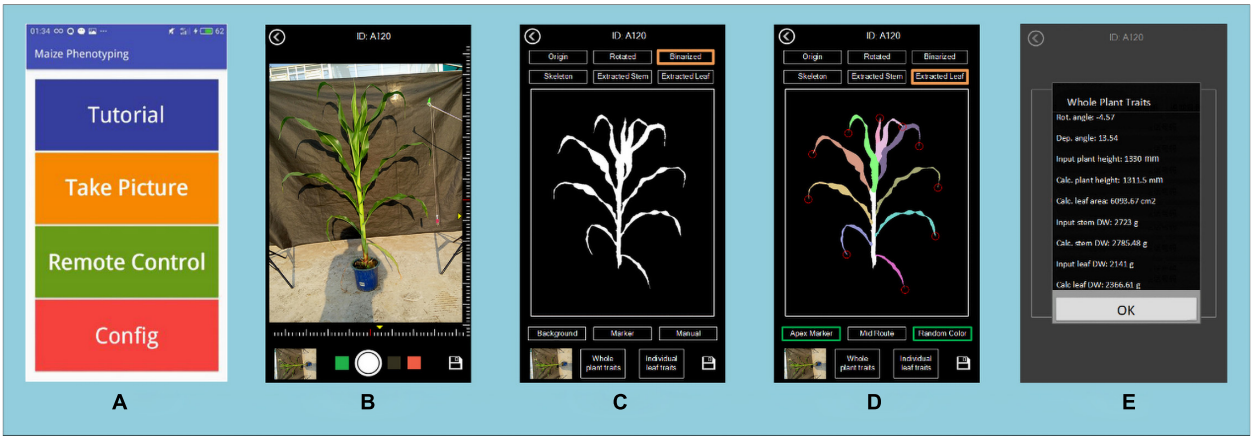


FIGURE 6 | The user interface of PocketMaize showing **(A)** menu; **(B)** image capturing page; **(C)** segmentation result page; **(D)** leaf extraction result page; **(E)** traits extraction result page.

Our own dataset included 720 images in the training set and 80 images in the validation set. A horizontal flip is applied to each image to produce a final training set of 1,440 training and 190 validation images. The transfer training was started with an initialized model pre-trained on the VOC 2012 dataset. The loss weight of the loss function is modified according to the total pixel size of background and the plant. The logit layer and the last layer are excluded to train on our own dataset.

Thinning Algorithm

The media axis of the segmented image is essential for stem and leaf recognition and the calculation of traits such as stem height and leaf length. Thinning/peeling-based methods such as Zhang's thinning algorithm (Zhang and Suen, 1984) will produce numerous spurs and are time-consuming, and Voronoi diagram-based methods will have difficulty deciding whether a skeletal branch should be pruned. Since the plant's binary images in this study are relatively large (originally up to 20 million pixels and will become even larger after angle calibration in this manuscript and might be larger for more advanced smartphones) and the boundary of the plant is usually very complicated and meandering, a proper way is to use distance transform-based methods. We developed a two-step skeletonization algorithm (**Figure 4**) based on the distance transform algorithm (Felzenszwalb and Huttenlocher, 2004). First, a distance transform algorithm was applied to the binary image. Define S_t to be the point set of the target skeleton we needed and S_n to be the point set containing all the points whose value in the distance transformed image is larger than at least n points in its eight neighbors. **Figure 4A** displays the original segmented image, and **Figures 4B,C** show the points in S_2 and S_3. We have approximately $S_2 \subseteq S_t \subseteq S_3$. In the second step, we designed a path finding algorithm to find a way to connect the points in S_2 with the points in S_3. **Figure 4D** is the result of our algorithm with the origin points in S_2 marks in blue and the connected pixels from S_3 marks in red.

Stem and Leaf Extraction

The stem was extracted by finding the shared route connecting the upper part and the plant root, and **Figure 5** presents the practical steps. First, all the endpoints of the skeleton image are detected. **Figure 5A** is the image of the mid axis and all the endpoints. Then, the shortest routes between the lowest endpoint and each of the other endpoints are traced and summed. **Figures 5B–E** shows this tracing procedure from lower leaves to higher leaves, where each individual leaf is marked in different colors and the overlaid route is marked in red.

Leaf apexes were located at the endpoints of the skeleton image. The leaf direction can be traced by finding the shortest route between leaf apexes and plant stems along the skeleton. **Figure 5E** displays the segmented plant stem and individual leaves painted in different colors. The leaf insertion angle and the leaf angles from the insertion to the tip can be directly measured from the leaf mid axis. With the stem area removed, the leaves in the lower half were naturally separated.

Traits Extraction

Finally, from the segmented images and skeletonized images, we calculated 45 traits (**Table 1**), which included 15 plant traits, 25 leaf traits and 5 stem traits. We can also extract each individual leaf and analyze the difference between leaves at the upper part of the plant and the lower part.

RESULTS

Development of a Smartphone Application: PocketMaize

In this study, equipped with an Android smartphone (Meizu MX4, MediaTek6595, CPU A17 2.2 GHz × 4 + A7 1.7 GHz × 4, GPU PowerVR G6200), the application was developed with two sensors: an RGB camera and a 3-axis accelerometer. In the image taking mode, an indicator displaying the current depression angle and rotation angle allows users to adjust phone orientation to

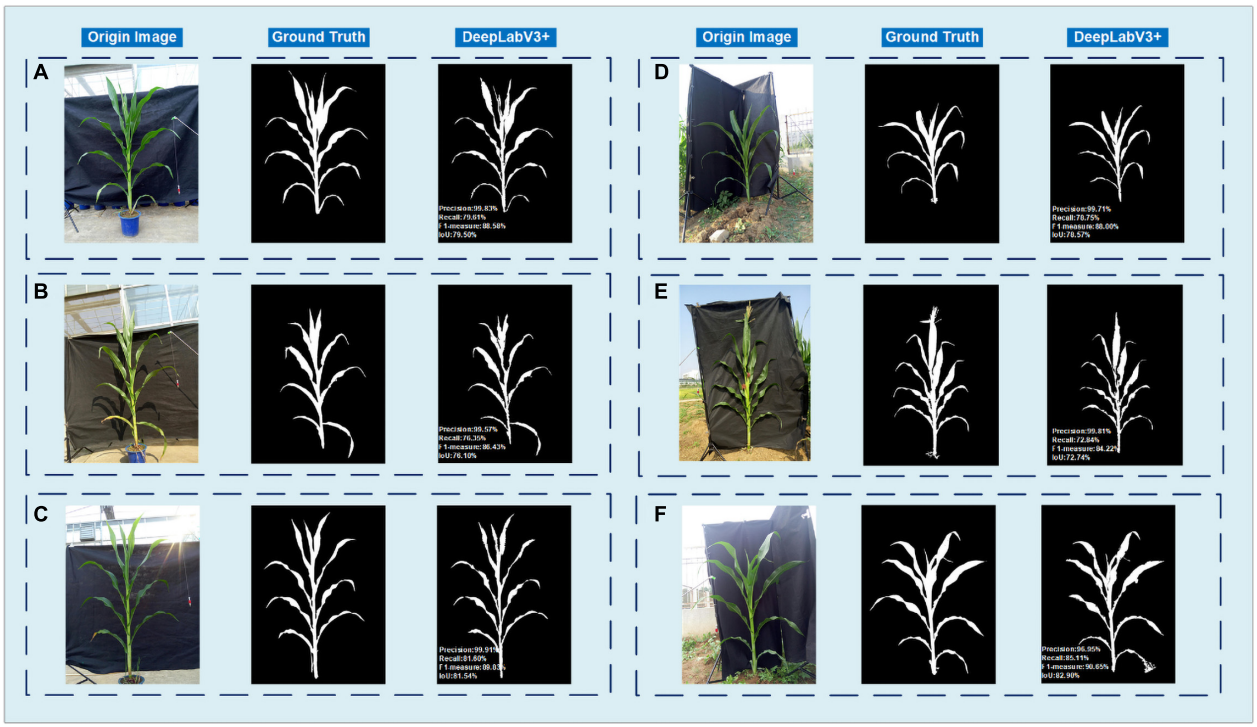


FIGURE 7 | Comparison of manual ground truth and the segmentation result of DeepLabV3+. The left three columns **(A–C)** are the results of three potted samples, and the right three columns **(D–F)** are the results of three in-field samples. The first row **(A,D)** shows samples taken in sunny mornings with even illumination. The second row **(B,E)** shows two samples taken at dawn when the images have a heavy yellow color deviation. The last row **(C,F)** is taken at midday with high brightness.

obtain the appropriate angles. Images were stored in JPG format with a resolution of 3,936 × 5,248 pixels. The camera was autofocused; ISO, shutter speed, and light balance were autofixed. Other necessary information, such as time, date, and plant ID, could also be manually input.

Image processing, processed images and extracted traits can be displayed and saved on the device. The final traits of the maize are stored in a CSV file. **Figure 6** shows the user interface of the application, which includes the main menu (A), the image taking page (B), the result of segmentation and stem and leaf recognition (C,D), and the traits displaying page (E).

Performance Evaluation of DeepLabV3+ Segmentation

In this study, after 1,440 images of maize were used to train the DeepLabV3+ model, another 190 images, including 95 images of potted samples and 95 in-field samples, were selected to test the DeepLabV3+ model. To evaluate the performance, four indicators, including precision, recall, F1-measure and IoU, are adopted. **Figure 7** shows the results of six samples under different conditions. The left three columns (A–C) are the results of three potted samples, and the right three columns (D–F) are the results of three in-field samples. The first row (A,D) shows samples taken in sunny mornings with even illumination. The second row (B,E) shows two samples taken at dawn when the images have a heavy yellow color deviation. The last row (C,F) is taken at midday with high brightness. In general, the DeepLabV3+ model works well in different color temperatures, different light intensities and mild wind or mild rainy days. However, a sun halo might influence the segmentation result. Although heavy wind will not affect the segmentation stage, it will decrease the accuracy in later stem and leaf recognition stage since the structure of the plant may greatly change.

For the DeepLabV3+ model, the mean values of the Precision, Recall, F1-measure and IoU are 97.31, 74.85, 86.10, and 79.91%, respectively.

Accuracy Evaluation of Plant Height Measurement

Plant height is the vertical distance from the bottom of the stem at soil surface to the top position of the while plant. To evaluate the measurement accuracy of plant height, all the plants were manually measured. The plant height was measured after the images were captured, and automatic plant height measurement was used to calculate the actual distance between the bottom position of the stem and the top of the whole plant. **Figure 8** shows the plant height results of manual observation versus automatic observation in all four depression degrees for all plants. The MAPE values were 3.556% for potted samples and 4.594% for field samples, and the R^2 coefficients were 0.928 and 0.958, respectively. The results show that smartphone applications

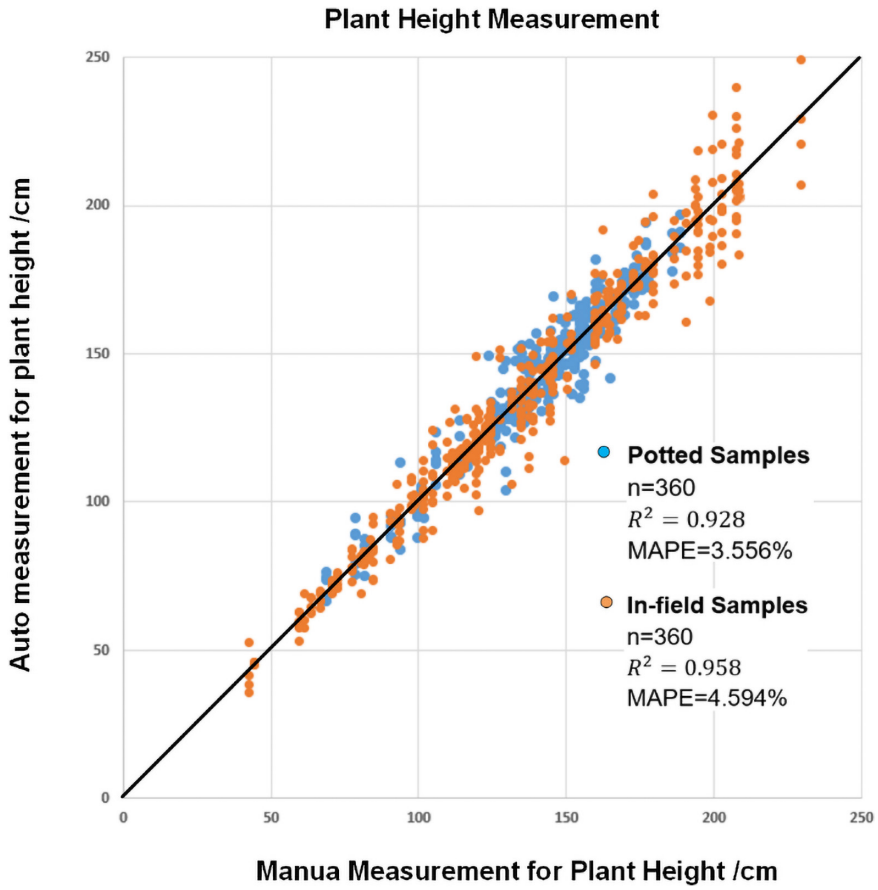


FIGURE 8 | The result of automatic plant height measurement versus manual plant height measurement.

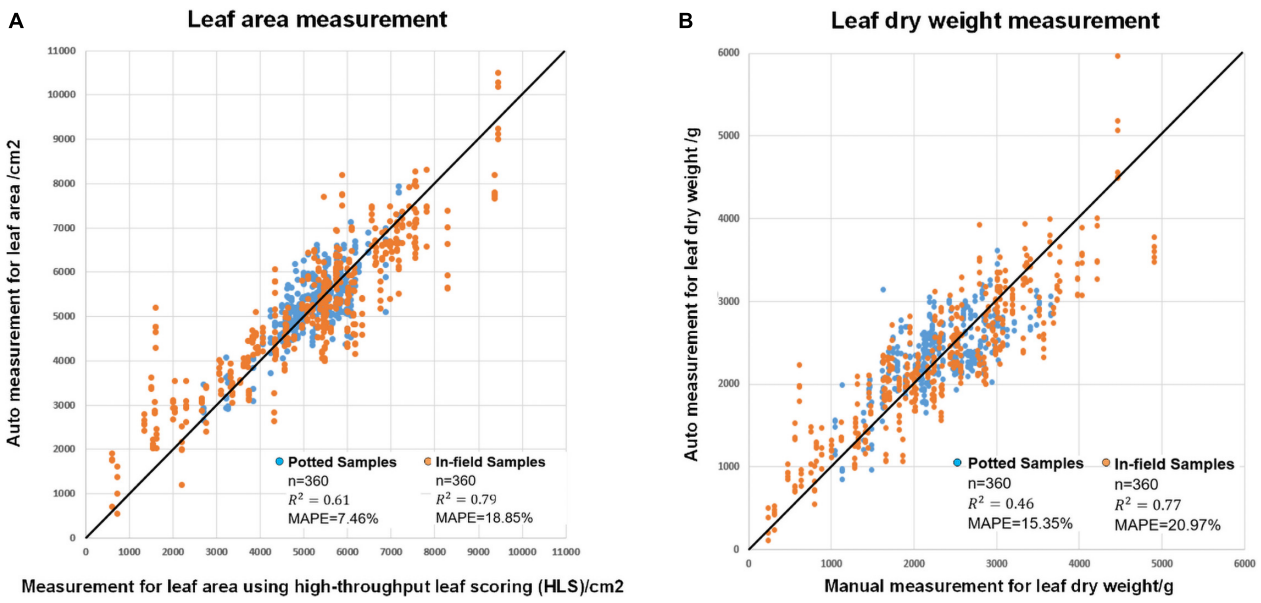


FIGURE 9 | The modeling result of **(A)** automatic leaf area measurement versus measurement using high-throughput leaf scoring (HLS) and **(B)** automatic leaf dry height measurement versus manual leaf dry height measurement.

FIGURE 10 | The modeling result for automatic stem dry weight measurement and manual stem dry weight measurement.

can correctly detect stems and have good potential for accurate measurement.

Accuracy Evaluation of Leaf Area and Leaf Dry Weight

Figure 9 shows the results of leaf area estimation (A) and leaf dry weight estimation (B). The MAPE values were 7.46% for potted leaf area, 18.85% for in-field leaf area, 15.35% for potted leaf dry weight and 20.97% for in-field leaf dry weight estimation. The squares of the correlation coefficients (R^2) were 0.61, 0.79, 0.46, and 0.77. The detailed model summaries for stepwise regression analysis for leaf area estimation and leaf dry weight estimation are shown in **Supplementary Tables 1, 2**.

Accuracy Evaluation of Stem Dry Weight

Since maize stems and maize leaves have a significant difference in organ structure and density, it is natural to evaluate the stem dry weight and leaf dry weight separately. In particular, the plant stem can be approximately seen as a cylinder, so the plant stem's dry weight can be estimated with the volume of a cylinder fitted to the stem together with other traits such as stem projected area, stem height and average stem width. **Figure 10** shows the result of stem biomass measurement. The MAPE values were 16.68 and 23.85% for potted and field samples, respectively. The squares of the correlation coefficients (R^2) were 0.64 and 0.88, respectively. The detailed model summaries for stepwise regression analysis for leaf area estimation and leaf dry weight estimation are shown in **Supplementary Table 3**.

DISCUSSION

Comparison of Trait Extraction With/Without Depression Angle Calibration

Since the difference in the depression angle can greatly change the original image, a depression angle calibration is essential before advanced image processing. **Figure 11** shows an example of eight images of one potted sample and one in-field sample with different depression angles and rotation angles. The actual depression angles/rotation angles are −2.1°/4.9°, 8.4°/5.7°, 15.2°/4.5°, and 22.7°/3.8° for the potted sample shown in A and 4.4°/3.68°, 12.1°/3.5°, 16.6°/2.89°, and 18.4°/3.6° for the field sample shown in B. The calibrated images show that our angle

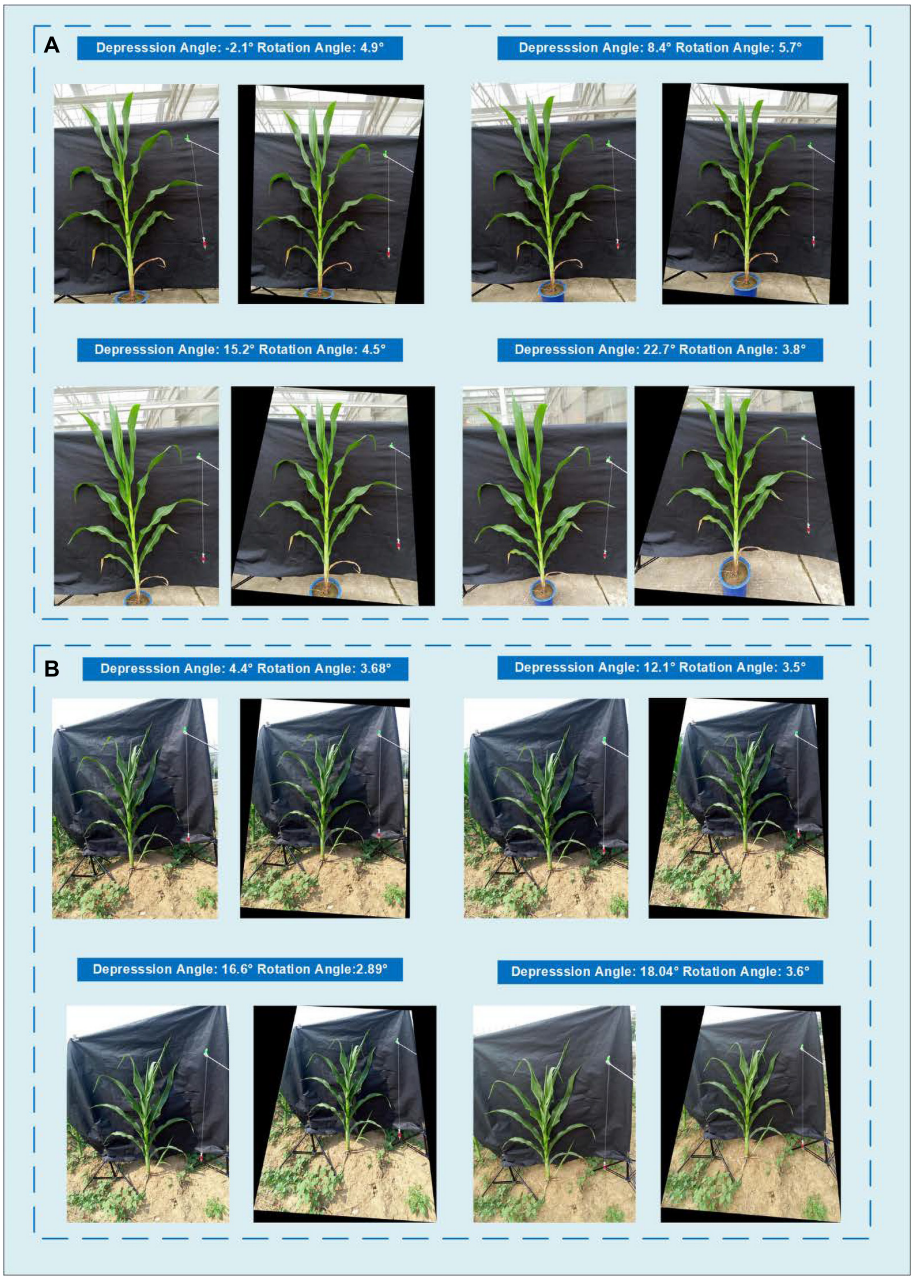


FIGURE 11 | Comparison of the images taken with different depression angles and their angle calibrated results showing **(A)** a potted sample and **(B)** an in-field sample.

calibration algorithm can vastly reduce the influence caused by different rotation and depression angles and transform the plant to an approximate front view.

Figure 12 shows the plant height, leaf area, leaf dry biomass and stem dry biomass with and without depression angle calibration. The results indicate that for plant height, leaf area and leaf dry weight, the result is still meaningful without depression calibration, with R^2 values up to 0.73, 0.65, and 0.57 for in-field samples (**Figures 12B,D,F**). However, a depression angle calibration can increase the measurement accuracy as R^2 increases to 0.99, 0.79, and 0.68 (**Figures 12A,C,E**). Stem dry weight can only be measured with depression angle calibration.

Comparison of Four Skeletonization Methods

The skeleton algorithm we developed is based on the distance transform algorithm. Several existing skeleton algorithms were tested during our application development, and some were

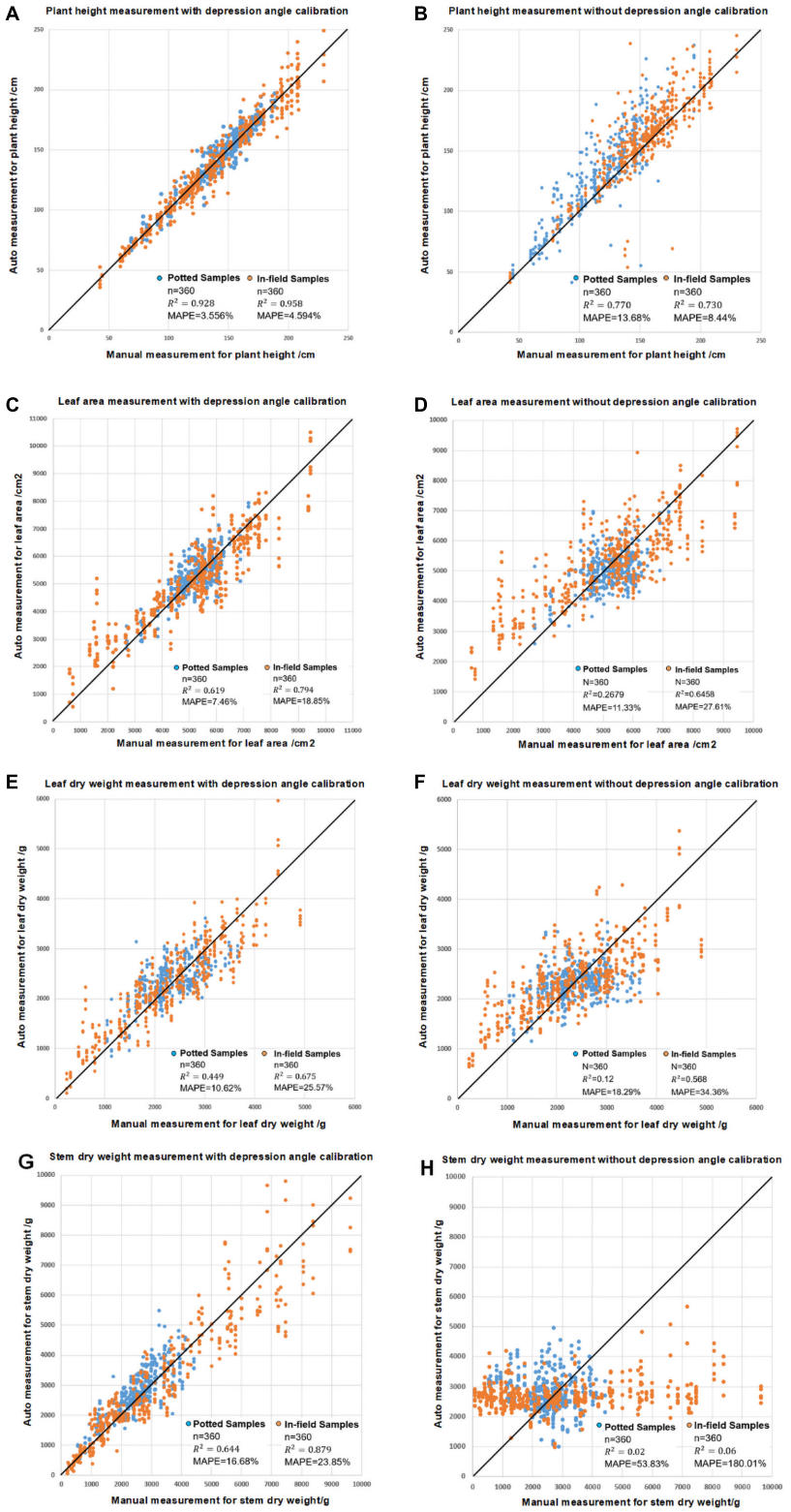


FIGURE 12 | The comparison of the modeling results for leaf area, leaf dry weight and stem dry weight measurements with and without angle calibration showing **(A)** plant height measurements with depression angle calibration, **(B)** plant height measurements without depression angle calibration, **(C)** leaf area measurements with calibration, **(D)** leaf area measurements without calibration, **(E)** leaf dry weight measurements with calibration, **(F)** leaf dry weight measurements without calibration, **(G)** stem dry weight measurements with calibration and **(H)** stem dry weight measurements without calibration.

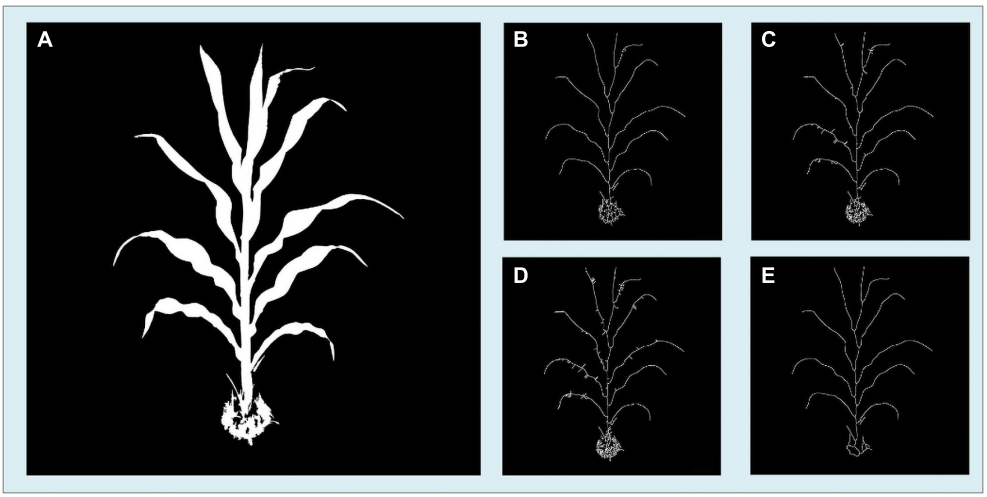


FIGURE 13 | Comparison of the skeletonization results of the four methods showing **(A)** the segmented image. **(B)** Result image of our algorithm. **(C)** Results image of Zhang's thinning algorithm with branch pruning. **(D)** Result image of Scikit-image's skeletonization. **(E)** Results image of heat equation-based skeletonization.

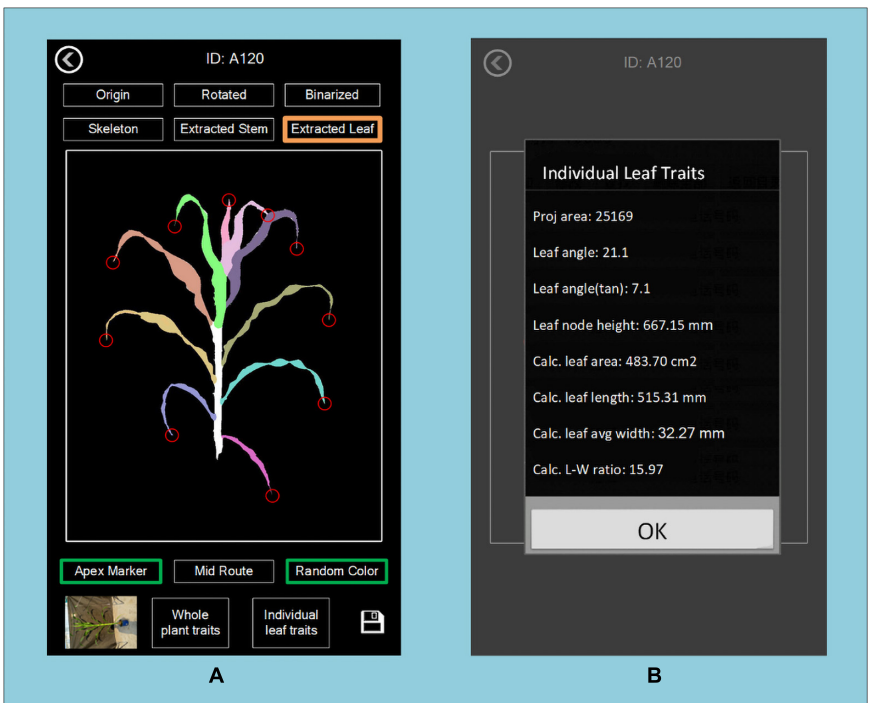


FIGURE 14 | Display of individual leaf traits at different heights of the plant showing **(A)** the whole plant and **(B)** traits of a specific leaf.

modified to match the situation better. It turned out that our algorithm has a good result both for correctness and calculation speed compared with the other algorithms. Our algorithm requires a shorter calculation time to find a maize plant's skeleton, yields fewer unexpected branches and burrs, and the skeleton is located closer to the center axis. In **Figure 13**, we present our algorithm's results compared with several other skeletonization algorithms. These candidate algorithms include Zhang's thinning algorithm (Zhang and Suen, 1984), the media axis algorithm provided by scikit-image (van der Walt et al., 2014), and 2D skeleton extraction based on the heat equation (Gao et al., 2018). **Figure 13A** is the original segmentation image, and **Figure 13B** is the result of our skeleton algorithm. **Figures 13C–E** are the skeletonization results of Zhang's thinning algorithm with branch pruning, the scikit-image's media axis algorithm, and the heat equation 2D skeleton extraction, respectively.

The details of these skeleton images show that our method's result has fewer unexpected branches, and the skeleton lies closer

to the center axis. Although our method does not have the best result among these four methods, it has the highest efficiency. **Supplementary Table 4** shows the calculation time consumption and memory use of our method and other methods. Our method is the fastest among these algorithms but consumes the largest amount of memory. With the development of smartphone chips, the processing memory of smartphones has become significantly larger. Therefore, our approach is a better choice for the skeletonization process on the smartphone platform.

Efficiency of the Image Process Procedure

In general, the average execution time for a single plant image of 20 million pixels is 2,482 ms operating on an Android smartphone (Meizu MX4, MediaTek6595,CPU A17 2.2 GHz × 4 + A7 1.7 GHz × 4,GPU PowerVR G6200) All image process-related algorithms were developed using C++ language combined with the OpenCV library and compiled into a Java library for Android. The image processing procedure contains three major parts: segmentation, skeletonization and trait calculation. The average computational times are 1,050, 641, and 791 ms for segmentation, skeletonization and trait calculation, respectively. The total computation time for the whole procedure varied from 700 to 4,000 ms depending on the complexity of the plant structure and cleanliness of the background. Moreover, a faster segmentation algorithm that can reduce the process time to less than 100 ms is provided in the application for clean backgrounds with stable environments.

Individual Leaf Traits Extraction

PocketMaize provides an algorithm to extract all individual leaves from one maize plant and to obtain the traits of each leaf. It provides data for evaluating the difference between leaves at higher places and lower places. As shown in **Figure 14**, traits of individual leaves can be examined and stored for further analysis for canopy research and to investigate leaf overlap and sunlight absorption at different layers of the plant.

Potential Application and Outlooks

Although the main object of this manuscript is to obtain single plant traits with high accuracy, the application can also calculate traits of several plants with minor overlapping, but the segmentation and traits calculation accuracy will decrease for severe overlapping. Moreover, the current work mainly focus on maize stem traits and leaf traits, the application to extract tassel traits and cob traits in reproductive stage will be improved in the future work. With training of enough images containing maize cobs and tassels, new segmentation model will be developed to obtain cobs traits during reproductive stage and estimate the final yield.

CONCLUSION

In conclusion, we developed PocketMaize, an android smartphone application for maize plant phenotyping. The application is capable of field and potted maize phenotyping without many additional devices used. A total of 45 traits, which included 15 plant traits, 25 leaf traits and 5 stem traits, were nondestructively extracted. The average execution time for a single plant image of 20 million pixels was within 3,500 ms. In the future, with more trained images, a portable and cost-effective phenotyping solution could be extended to maize functional genomics studies, maize breeding, and disease and insect pest detection.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

LL and WY designed the research, performed the experiments, analyzed the data, and wrote the manuscript. LY, DW, JY, and HF helped to perform the experiments. QL and WY supervised the project and helped to design the research. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Key Research and Development Program (2020YFD1000904-1-3), the National Natural Science Foundation of China (31770397), Major Science and Technology Projects in Hubei Province, Fundamental Research Funds for the Central Universities (2662020ZKPY017 and 2021ZKPY006), and Cooperative Funding between Huazhong Agricultural University and Shenzhen Institute of Agricultural Genomics (SZYJY2021005 and SZYJY2021007).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.770217/full#supplementary-material

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.* Available Online at: https://arxiv.org/abs/1603.04467 (accessed March 14, 2016).

Araus, J. L., and Kefauver, S. C. (2018). Breeding to adapt agriculture to climate change: affordable phenotyping solutions. *Curr. Opin. Plant Biol.* 45, 237–247. doi: 10.1016/j.pbi.2018.05.003

Awlia, M., Nigro, A., Fajkus, J., Schmoeckel, S. M., Negrão, S., Santelia, D., et al. (2016). High-throughput non-destructive phenotyping of traits that contribute

to salinity tolerance in *Arabidopsis thaliana*. *Front. Plant Sci.* 7:1414. doi: 10.3389/fpls.2016.01414

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., et al. (2016). Global conservation priorities for crop wild relatives. *Nat. Plants* 2:16022. doi: 10.1038/nplants.2016.22

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2016). DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *ArXiv* [Preprint]. Available Online at: https://arxiv.org/abs/1606.00915 (accessed June 2, 2016).

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. *ArXiv* [Preprint]. Available Online at: https://arxiv.org/abs/1606.00915v2 (accessed May 12, 2017).

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-Decoder with atrous separable convolution for semantic image segmentation. *ArXiv* [Preprint]. Available Online at: https://arxiv.org/abs/1802.02611v3 (accessed August 22, 2018).

Chollet, F. (2017). Xception: deep learning with depthwise separable convolutions. *ArXiv* [Preprint]. Available Online at: https://arxiv.org/abs/1610.02357 (accessed April 4, 2017).

Chung, Y. S., Choi, S. C., Silva, R. R., Kang, J. W., Eom, J. H., and Kim, C. (2017). Case study: estimation of sorghum biomass using digital image analysis with Canopeo. *Biomass Bioenergy* 105, 207–210. doi: 10.1016/j.biombioe.2017.06.027

Confalonieri, R., Paleari, L., Foi, M., Movedi, E., Vesely, F. M., Thoelke, W., et al. (2017). PocketPlant3D: analysing canopy structure using a smartphone. *Biosyst. Eng.* 164, 1–12. doi: 10.1016/j.biosystemseng.2017.09.014

Confalonieri, R., Paleari, L., Movedi, E., Pagani, V., Orlando, F., Foi, M., et al. (2015). Improving in vivo plant nitrogen content estimates from digital images: trueness and precision of a new approach as compared to other methods and commercial devices. *Biosyst. Eng.* 135, 21–30. doi: 10.1016/j.biosystemseng.2015.04.013

Das Choudhury, S., Stoerger, V., Samal, A., Schnable, J. C., Liang, Z., and Yu, J. G. (2016). "Automated vegetative stage phenotyping analysis of maize plants using visible light images," in *KDD workshop on Data Science for Food, Energy and Water (KDD-DSFEW)*, San Francisco: Association for Computing Machinery (ACM).

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082. doi: 10.2135/cropsci2016.10.0885

Efroni, I., Eshed, Y., and Lifschitz, E. (2010). Morphogenesis of simple and compound leaves: a critical review. *Plant Cell* 22, 1019–1032. doi: 10.1105/tpc.109.073601

Fang, W., Feng, H., Yang, W., Duan, L., Chen, G., Xiong, L., et al. (2016). High-throughput volumetric reconstruction for 3D wheat plant architecture studies. *J. Innov. Opt. Health Sci.* 09:1650037. doi: 10.1142/S1793545816500371

FAO (2010). *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. Rome: FAO.

Felzenszwalb, P. F., and Huttenlocher, D. P. (2004). Distance transforms of sampled functions. *Theory Comput.* 8, 415–428.

Feng, H., Jiang, N., Huang, C., Fang, W., Yang, W., Chen, G., et al. (2013). A hyperspectral imaging system for an accurate prediction of the above-ground biomass of individual rice plants. *Rev. Sci. Instrum.* 84:095107. doi: 10.1063/1.4818918

Gao, F., Wei, G., Xin, S., Gao, S., and Zhou, Y. (2018). 2D skeleton extraction based on heat equation. *Comput. Graph.* 74, 99–108. doi: 10.1016/j.cag.2018.05.005

Golzarian, M. R., Frick, R. A., Rajendran, K., Berger, B., Roy, S., Tester, M., et al. (2011). Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* 7:2. doi: 10.1186/1746-4811-7-2

Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., et al. (2006). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.* 169, 623–635. doi: 10.1111/j.1469-8137.2005.01609.x

Hallau, L., Neumann, M., Klatt, B., Kleinhenz, B., Klein, T., Kuhn, C., et al. (2018). Automated identification of sugar beet diseases using smartphones. *Plant Pathol.* 67, 399–410. doi: 10.1111/ppa.12741

He, K., Zhang, X., Ren, S., and Sun, J. (2015). *Deep Residual Learning for Image Recognition*. Manhattan: IEEE.

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]. Available Online at: https://arxiv.org/abs/1704.04861 (accessed April 17, 2017).

Jansen, M., Gilmer, F., Biskup, B., Nagel, K. A., Rascher, U., Fischbach, A., et al. (2009). Simultaneous phenotyping of leaf growth and chlorophyll fluorescence *via* GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Funct. Plant Biol.* 36, 902–914. doi: 10.1071/FP09095

Kaneda, Y., Shibata, S., and Mineno, H. (2017). Multi-modal sliding window-based support vector regression for predicting plant water stress. *Knowl. Based Syst.* 134, 135–148. doi: 10.1016/j.knosys.2017.07.028

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., and Mardis, E. R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell* 155, 27–38. doi: 10.1016/j.cell.2013.09.006

Liu, Z., Wang, S., and Ding, X. (2012). "ROI perspective transform based road marking detection and recognition," in *2012 International Conference on Audio, Language and Image Processing*, (Manhattan: IEEE), 841–846.

Orlando, F., Movedi, E., Coduto, D., Parisi, S., Brancadoro, L., Pagani, V., et al. (2016). Estimating Leaf Area Index (LAI) in vineyards using the PocketLAI smart-app. *Sensors* 16:2004. doi: 10.3390/s16122004

Patrignani, A., and Ochsner, T. E. (2015). Canopeo: a powerful new tool for measuring fractional green canopy cover. *Agron. J.* 107, 2312–2320. doi: 10.2134/agronj15.0150

Pethybridge, S. J., and Nelson, S. C. (2018). Estimate, a New iPad application for assessment of plant disease severity using photographic standard area diagrams. *Plant Dis.* 102, 276–281. doi: 10.1094/PDIS-07-17-1094-SR

Pilling, D., Bélanger, J., Diulgheroff, S., Koskela, J., Leroy, G., Mair, G., et al. (2020). Global status of genetic resources for food and agriculture: challenges and research needs: global status of genetic resources for food and agriculture. *Genet. Resour.* 1, 4–16. doi: 10.org/10.46265/genresj.2020.1.4-16

Ritchie, H., and Roser, M. (2020). *Agricultural Production: Our World in Data*. Oxford: University of Oxford.

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

Tisné, S., Serrand, Y., Bach, L., Gilbault, E., Ben Ameur, R., Balasse, H., et al. (2013). Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. *Plant J.* 74, 534–544. doi: 10.1111/tpj.12131

Torii, T. (2000). Research in autonomous agriculture vehicles in Japan. *Comput. Electron. Agric.* 25, 133–153. doi: 10.org/10.1016/S0168-1699(99)00060-5

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.org/10.7717/peerj.453

Walter, A., Scharr, H., Gilmer, F., Zierer, R., Nagel, K. A., Ernst, M., et al. (2007). Dynamics of seedling growth acclimation towards altered light conditions can be quantified *via* GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytol.* 174, 447–455. doi: 10.1111/j.1469-8137.2007.02002.x

Wu, D., Guo, Z., Ye, J., Feng, H., Liu, J., Chen, G., et al. (2019). Combining high-throughput micro-CT-RGB phenotyping and genome-wide association study to dissect the genetic architecture of tiller growth in rice. *J. Exp. Bot.* 70, 545–561. doi: 10.1093/jxb/ery373

Yang, W., Duan, L., Chen, G., Xiong, L., and Liu, Q. (2013). Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. *Curr. Opin. Plant Biol.* 16, 180–187. doi: 10.1016/j.pbi.2013.03.005

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, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087

Yang, W., Guo, Z., Huang, C., Wang, K., Jiang, N., Feng, H., et al. (2015). Genome-wide association study of rice (*Oryza sativa* L.) leaf traits with a high-throughput leaf scorer. *J. Exp. Bot.* 66, 5605–5615. doi: 10.1093/jxb/erv100

Zhang, T. Y., and Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Commun. ACM* 27, 236–239.

Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., et al. (2017). High-Throughput Phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiol.* 173, 1554–1564. doi: 10.1104/pp.16.01516

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 1330–1334. doi: 10.1109/34.888718

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fpls.2021.792244

Diseases Detection of Occlusion and Overlapping Tomato Leaves Based on Deep Learning

Xuewei Wang[1], Jun Liu[1] and Guoxu Liu[2]*

[1] Shandong Provincial University Laboratory for Protected Horticulture, Blockchain Laboratory of Agricultural Vegetables, Weifang University of Science and Technology, Weifang, China, [2] College of Information and Control Engineering, Weifang University, Weifang, China

Background: In view of the existence of light shadow, branches occlusion, and leaves overlapping conditions in the real natural environment, problems such as slow detection speed, low detection accuracy, high missed detection rate, and poor robustness in plant diseases and pests detection technology arise.

Results: Based on YOLOv3-tiny network architecture, to reduce layer-by-layer loss of information during network transmission, and to learn from the idea of inverse-residual block, this study proposes a YOLOv3-tiny-IRB algorithm to optimize its feature extraction network, improve the gradient disappearance phenomenon during network deepening, avoid feature information loss, and realize network multilayer feature multiplexing and fusion. The network is trained by the methods of expanding datasets and multiscale strategies to obtain the optimal weight model.

Conclusion: The experimental results show that when the method is tested on the self-built tomato diseases and pests dataset, and while ensuring the detection speed (206 frame rate per second), the mean Average precision (mAP) under three conditions: (a) deep separation, (b) debris occlusion, and (c) leaves overlapping are 98.3, 92.1, and 90.2%, respectively. Compared with the current mainstream object detection methods, the proposed method improves the detection accuracy of tomato diseases and pests under conditions of occlusion and overlapping in real natural environment.

Keywords: YOLOv3-tiny, inverse-residual block, field images, multi-scale, occlusion and overlapping, robust

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Dawid Polap,
Silesian University of Technology, Poland
Neetu Mittal,
Amity University, India
Chengcheng Chen,
Jilin University, China

****Correspondence:***
Jun Liu
liu_jun860116@wfust.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *10 October 2021*
Accepted: *09 November 2021*
Published: *10 December 2021*

Citation:
Wang X, Liu J and Liu G (2021) Diseases Detection of Occlusion and Overlapping Tomato Leaves Based on Deep Learning. Front. Plant Sci. 12:792244. doi: 10.3389/fpls.2021.792244

INTRODUCTION

Tomato is one of the most popular crops planted in China, and it has an irreplaceable position in vegetables, fruits, medicinal, and other aspects, with a huge planting volume and demand (Li, 2012). Taking Shouguang City, Shandong Province as an example, Shouguang City's tomatoes are mainly produced in Luocheng Street, with about 12,000 household greenhouse planters and 20,000 winter-warm greenhouses, with an annual trading volume of 360 million kilograms, and an annual trading volume of 730 million yuan. The products are exported to Russia, North Korea, Myanmar, and other countries. This town is an important tomato production and sales base in Shandong Province and enjoys the reputation of "small town with tomato characteristics." According to statistics, a common 100-m greenhouse has a revenue of at least RMB 100,000, and vegetable farmers have realized the "income-increasing dream" through tomatoes.

Traditionally speaking, tomatoes belong to seasonal fruits and vegetables, but the market has a great demand for tomatoes in each season. To meet the market demand and improve economic benefits, most farmers use greenhouse planting to overcome the influence of season, temperature, and other environmental factors, and achieve tomato planting and production for more than three seasons in 1 year. From the previous field research and feedback from farmers, we know that the whole growth cycle of tomatoes has strict requirements on the growth environment, planting methods, pest control, and other aspects, and the requirements on the external environment of tomatoes at all growth stages of the whole growth cycle are also of high standards and are different. In recent years, the impact of diseases and pests on tomato cultivation has been aggravated. The main reason is that the optimized planting structure and complete water and fertilizer supply conditions are not only conducive for tomato growth, but also provide convenience for the occurrence of diseases and pests. At the same time, the unscientific and non-standard use of pesticides also cause the increasing resistance of pathogens. Also, the differences of tomato varieties and cross-hazards are the causes of the growing severity of tomato diseases and pests (Wang et al., 2018). Therefore, the cost, time, and labor consumption of high-quality tomato cultivation are relatively high. However, most of the peasant households have not received professional knowledge, and they do not know the symptoms of diseases, pests, and other causative factors.

During tomato planting, the information of diseases and pests, the demand of crop growth environment, and the control measures mostly depend on the communication between peasant households and the previous planting experience. It is difficult to grasp the diseases and pests that may occur in a certain planting stage under certain conditions. It is also difficult to accurately determine the types of diseases and pests and their control methods. These practical problems have a great impact on tomato production.

The investigation revealed that the major diseases of tomato included 34 infestation diseases and 39 physiological diseases, and the disease characteristics were mainly focused on the color and morphology of the lesions. To make the research typical and better feasible, in this work, 12 common diseases including early blight, late blight, yellow leaf curl virus, gray leaf, coal pollution, gray mold, leaf mold, navel rot, leaf curl disease, mosaic, leaf miner, and greenhouse whitefly were selected for research.

The traditional method of identifying tomato diseases and pests is usually manual identification, that is, growers make subjective judgment based on planting experience or text data, or image comparison through the network, books, etc., or ask pathologists to analyze and identify tomato diseases and pests. Traditional manual diseases and pests identification takes a lot of time and effort, and is often accompanied by very high subjectivity. Subjective evaluation is susceptible to personal factors and external factors (such as light, occlusion, and overlapping). It is inefficient and has large errors, which can easily lead to the wrong diagnosis and wrong medication of tomato diseases and pests. Severe conditions can also cause pollution to water sources, soil, and so on. In addition, due to the scattered agricultural production in China and the lack of relevant agricultural experts, there are some limitations in the support provided by tomato pathologists in the professional pathological analysis and decision methods. Therefore, the manual identification of diseases and pests cannot meet the requirements of high-efficiency tomato production in the development of modern agriculture, an so the automatic and accurate detection of tomato diseases and pests is urgently needed.

Traditional plant disease detection relies on a large amount of manual design, where the model generalization performance is poor and the detection accuracy cannot meet the practical demand. Thanks to the rapid development of deep learning, Girshick et al. (2013) proposed that Region-CNN (R-CNN) and the precision of detection was substantially improved. Subsequent researchers have made improvements from a number of perspectives based on R-CNN. Fast R-CNN (Girshick, 2015) was proposed, and the detection efficiency is improved by sharing the multitask loss function and convolution weights. Faster R-CNN (Ren et al., 2017) integrates region nominations with convolutional neural networks and truly implements an end-to-end target detection framework. Mask R-CNN (He et al., 2017) introduced region of interest (ROI) align to replace ROI pooling and enable segmentation and detection of images. Region-based Fully Convolutional Networks (R-FCN) (Dai et al., 2016) introduced fully convolutional operation and the detection effect is improved greatly. FPN (Lin et al., 2017) and CascadeR-CNN (Cai and Vasconcelos, 2017) have achieved an extremely high detection accuracy and approximate the resolving power of humans. The above detection framework all contain both regional nominations and detection networks, and they are called two-stage methods. Other researchers have proposed region free nomination stage that unifies classification and detection tasks, and they are called one-stage methods. For example, YOLO (Redmon et al., 2016), RetinaNet (Lin et al., 2018), RefineDet (Zhang et al., 2017) are typical one-stage methods, and the real-time performance is greatly improved.

With the deep integration of modern information technology such as Internet of Things, Cloud Computing, and Artificial Intelligence with agriculture, smart agriculture has become a major trend in the development of modern agriculture in the world through the implementation of whole industry supply chain with real-time information perception, quantitative decision-making, intelligent production control, and precise personality management, and has made important progress in the field of crop harvesting (Xu et al., 2020). In the field of crop diseases and pests identification, Intelligent Agriculture relies on the Internet of Things system built by fixed monitoring cameras, mobile equipment, robots, smartphones, and other terminals. The classification and detection of network based on deep learning method is studied on the basis of real-time collection and acquisition of a large number of high-quality image dataset of crop diseases and pests, which can provide accurate, low-cost, high-efficiency, reliable, and real-time results for broad agricultural producers. It has gradually become the focus of research at home and abroad. Computer vision provides a very effective means for automatic detection of crop diseases and pests, and some progress has been made (Ouhami et al.,

2020). Under natural scenes, the tomato diseases and pest objects are often covered by light and shade, and the branches and leaves are covered or in an overlapping state. The identification and localization of the tomato diseases and pests objects under the influence of shading and overlapping is a difficult problem that must be solved.

With the rapid development of smart agriculture, the technology of using cameras to determine whether plants appear in images are infected with diseases and pests has been applied in the field of smart agriculture, which plays an increasingly important role in plant protection. This technique of using computer vision and machine learning to determine whether a particular plant in a camera is affected by diseases or pests is called plant diseases and pests identification, as shown in **Figure 1**.

Plant diseases and pests identification not only has a very urgent application need, but also has a very important research value. In recent years, plant diseases and pest identification has attracted wide attention from academia and agriculture, and become a research hotspot in the field of computer vision. After more than 10 years of development, a large number of plant diseases and pests identification models have been proposed at home and abroad, and very high accuracy has been achieved under the limited simulation conditions (Singh et al., 2018; Geetharamani and Arun Pandian, 2019; Shekhawat and Sinha, 2020), and even surpasses the ability of human vision.

In recent years, some progress has been made in the research of plant diseases and pest identification under natural scenes. Fuentes et al. (2019) proposed an improved Faster R-CNN algorithm, which can effectively detect and locate plant abnormalities. The average accuracy of 92.5% was achieved in the built tomato plant abnormality description dataset. However, the real-time performance of the algorithm was not strong. Anagnostis et al. (2020) took images of walnut anthrax in orchards under various light conditions. A total of 4,491 images of leaves with and without anthrax were collected. The images of leaves infected with anthrax reached 2,356, slightly more than the images of healthy leaves. The classification accuracy of walnut anthrax was as high as 98.719% using convolutional neural network. Prabhakar et al. (2020) used ResNet101 to measure the severity of early blight of tomato leaves and the accuracy reached 94.6%. But their methods can only judge whether the disease was infected or not and cannot locate the disease. They mainly aim at the target recognition of a small number of images taken in close range, which is difficult to apply in practice. Zhao and Qu (2019a,b) used YOLOv2 algorithm to detect healthy and diseased tomatoes, and the mean Average precision (mAP) was as high as 91%. However, the method did not take into account the small and dense multiobject occlusion and overlap problem in natural environment. Liu and Wang (2020a,b) proposed an improved YOLO V3 algorithm for tomato diseases and pests detection with high accuracy and speed, but this method can only effectively detect tomato diseases and pests targets in the case of slight leaf overlap, and there is no satisfactory detection result in the case of large area occlusion. There are often uncertainty issues such as posture, background, and occlusion of leaves in the detection of plant diseases, which can greatly affect the detection accuracy. So, increasing the object detection accuracy has always been paid great attention to Liu et al. (2020) proposed an improved tomato detection model based on YOLOv3 aiming at complex environmental conditions, such as light change, branching, leaf blockage, and tomato overlap, which integrated a dense architecture for feature reuse, but the model was only used for tomato fruit positioning and could not be used for tomato diseases and pests detection. All the above literatures utilize the excellent learning ability, flexibility, and adaptability of convolutional neural network to solve the problems of time-consuming, laborious, and low accuracy in plant diseases and pests detection under complex background. However, in the above studies, the leaves of plant diseases and pests are mostly sparse and complete, and the characteristics of diseases and pests are obvious. In this work, the images of tomato diseases and pests are collected under different light conditions in the real natural environment, and there are even sunlight shadows or sundries, such as branches and trunks, or the leaves overlap densely. These factors are obstacles in the detection of tomato diseases and pests. To effectively carry out real-time detection for multiple objects, an improved object detection model based on YOLOv3 needs to be proposed for issues such as small objects and occluded objects prone to being missed or inaccurate detection frame positioning.

In the real natural environment, the study of plant diseases and pests identification has its particularity. In the real agricultural Internet of Things video monitoring, there are various shooting equipments, the image quality of plant diseases and pests objects is poor, the resolution is low, and there are also obvious changes in perspective and light (Barbedo, 2018). Therefore, compared with general image recognition, plant diseases and pests identification still faces the following problems: (1) In different monitoring and shooting equipment, the distance between plants and shooting equipment is different, resulting in different resolution, light, and perspective of plant diseases and pests image under different shooting equipment horizons, and different visual characteristics of the same plant diseases and pests image will produce obvious changes; (2) Different degree of occlusion caused by background and other factors leads to a large number of occlusion problems, which lead to poor identification of plant diseases and pests; (3) Due to the changes of leaf posture and shooting equipment angle, the differences of visual characteristics between different images of plant diseases and pests may be small in different shooting devices. In addition, some specific problems have not been paid enough attention to. For example, large-scale and fast retrieval problems, insufficient data problems, complex and crossmodal problems of plant diseases and pests occurrence in the actual agricultural environment, make the problem of plant diseases and pests identification more difficult than the general case-based image retrieval.

Investigation on the field environment of tomato greenhouse base showed that tomato plants grew densely, and light shading, branch and leaf occlusion, and overlap accounted for about 21.2%. Thus, tomato diseases and pests detection under conditions of occlusion and overlapping become the key and difficult point of the research. To solve the problem of rapid and high-precision detection of tomato diseases and pests in real natural environment, this work proposes a method

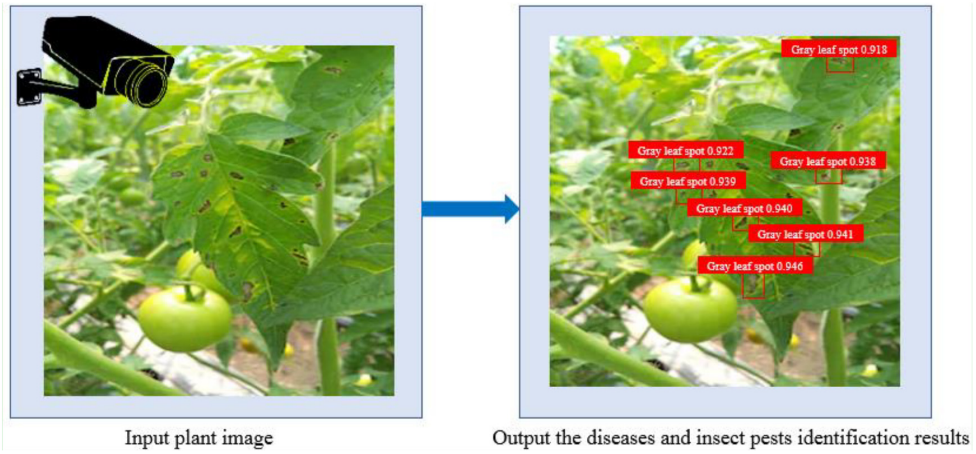


FIGURE 1 | An example illustrating plant diseases and pests identification.

to enhance the learning of foreground region features by occluding overlapping object foreground region samples, chooses YOLOv3-Tiny model based on regression method, and proposes a YOLOv3-tiny-IRB network structure with inverse residual blocks. Depth-wise convolution is used to reduce the model parameters, and an inverse residual module is constructed to extract high-dimensional features, and a linear activation function is used to reduce the loss of information caused by the channel combination process. The improved object detection network is trained by fusing data amplification and multiscale training strategies. The detection effect of the method in this study is significantly improved under two kinds of interference scenarios, i.e., sundries occlusion and blade overlap.

EXPERIMENTAL DATA

Dataset Acquisition

The experimental tomato planting base is located in Shouguang City, Shandong Province. By Using smartphone, digital camera, and other monitoring equipment with various resolutions, 15,000 images of early tomato diseases and pests during the growth and development period were collected. The weather during image acquisition includes sunny and cloudy days, and the acquisition period is 8:00-18:00, which covers possible lighting conditions such as sunshine, backlight, and sidelight. Greenhouse tomato leaves are photographed in multiple orientations so that the main features of the diseases and pests can be shot, such as texture, color, shape, etc. Each image is formatted as JPG. Images were cropped to 224 × 224 pixel size.

Five thousand images containing the following three representative scenarios were screened from 15,000 tomato diseases and pests images.

(a) Leaves sparse and complete. The objects are relatively clear and easy to identify.

(b) Branches or sunlight shade or other debris occlusion. It is possible that there are situations in which diseases and pests are too small, adherent, mutually obscured, or obscured by the shoot and leaves, increasing the difficulty of detection.

(c) Leaves overlapping densely. Overlapping bounding boxes may be erroneously discarded, leading to missed observations with a larger probability.

TABLE 1 | The number of each species of diseases and pests.

Species	Number
Early blight	401
Late blight	416
Gray leaf spot	425
Brown spot	431
Coal pollution	408
Gray mold	421
Leaf mold	419
Powdery mildew	402
Leaf curl	418
Mosaic	413
Leaf miner	411
Greenhouse whitefly	435
Total	5000

The number of each species of diseases and pests is shown in **Table 1**.

From 5,000 representative tomato diseases and pests images, 3,500 were randomly selected as original training images (containing 21,038 tomato diseases and pests objects), and the remaining 1,500 were selected as test images (containing 9,067 tomato diseases and pests objects).

Image Enhancement

The image enhancement of training samples can improve the quality and diversity of samples, which is conducive to the improvement of CNN detection accuracy (Ding and Taylor, 2016). Under natural light of greenhouse planting base, especially

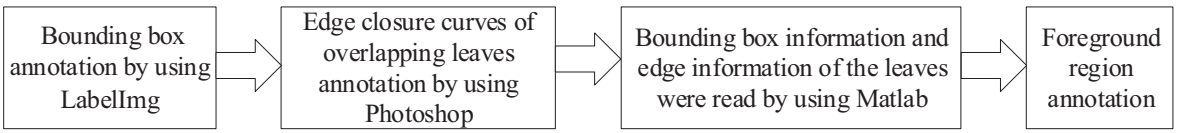


FIGURE 2 | The process of sample labeling.

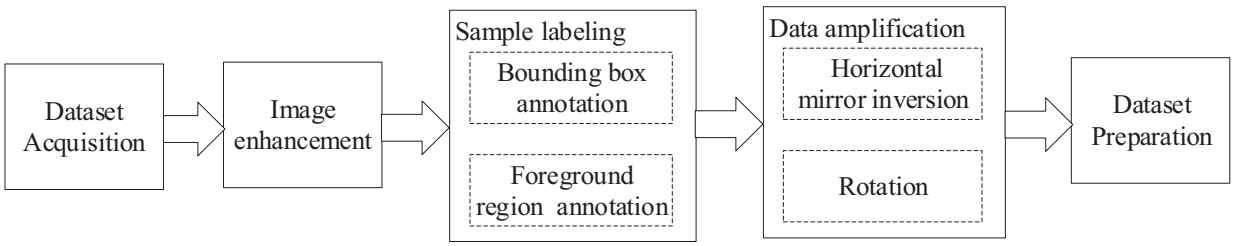


FIGURE 3 | The process of dataset preparation.

TABLE 2 | Datasets and sample size.

Datasets	Data processing method	Sample size	Number of annotation	
			Bounding box annotation	Foreground area annotation
A	No	3500	21038	2987
B	Image enhancement	3500	21038	2987
C	Data amplification	29016	173304	24158

when the light is very strong, due to the mutual occlusion of tomato plant leaves or backlight photography, the leaf surface produces shadows, which makes the image characteristics of tomato diseases and pests very different from those under normal light. Especially, some relatively small objects are not obvious in the image, affecting the quality of tomato diseases and pests image. The quality of training samples can affect the detection effect of the model, and so the contrast of the image needs to be adjusted to improve the detection effect of the detection model. In this study, adaptive histogram equalization (Algorithms, 2015) was used to enhance tomato diseases and pests images, improve the gray dynamic range of images, effectively improve the contrast of images, and enrich the details of images, which is equivalent to adjusting the image brightness and reducing the impact of light on image quality.

Sample Labeling

To improve the detection accuracy of tomato diseases and pests, various appearances and shapes of the objects were fully considered in the sample labeling process in this study. Manual labeling, interactive labeling, and Matlab programming were used for labeling. The process of sample labeling is shown in **Figure 2**.

(1) LabelImg, an open source annotation tool, was used to annotate the bounding boxes of 21,038 tomato diseases and pests objects in 3,500 original training images (no annotation was made when the object was covered by more than 70% of the area). Using this software, images in the dataset can be annotated as *.xml and *.txt files. The annotated file saves information such as class, size, and location of each object in the image. Also, LabelImg (TzuTa, 2017) was used to annotate 9,067 tomato diseases and pests object bounding box in 1,500 test images (no annotation was made when the object was covered by more than 70%). Considering that the test dataset is used to evaluate the detection accuracy of the model, the test dataset does not need to mark the object foreground area.

(2) The edge closure curves of overlapping leaves were automatically generated by using the quick selection tool of Photoshop software. However, for leaves with uneven surface color and illumination, it is difficult to automatically generate accurate edges, and the edge contours of leaves need to be manually marked.

(3) Using Matlab programming, the bounding box information and edge information of the leaves were read, and the pixels of the area outside the edge contour curve in the bounding box of the object were set to 0.

(4) In view of the difficulties caused by occlusion or overlap in tomato diseases and pests detection, a method of enhancing the learning of convolutional features of tomato diseases and pests foreground regions by annotating the foreground regions of training samples is proposed. Firstly, by manual annotation method, the pixels of the object background area were set to zero to obtain the object foreground area samples, and the object foreground area samples were trained to reduce the interference of non-foreground features in the bounding box, so as to enhance the learning of foreground features by the network and obtain the tomato diseases and pests detection network. When labeling samples, the pixels of the object background area in the labeling bounding box were set to 0, whereas the pixels of the foreground area remain unchanged. Thus, when convolution feature extraction is performed, the influence of unrelated information features on the feature extraction of

tomato diseases and pests can be reduced. At the same time, to retain the color, shape, and texture of the edge of tomato diseases and pests, the object foreground area in the bounding box included a certain area (5–10 pixels) around the object contour to enhance the model's learning of the characteristics of the foreground area including, the object edge.

Data Amplification

The training sample was expanded in this study. Considering that most of the tomato diseases and pests on the leaves of the tomato plant are naturally suspended, while some of them are inclined at multiple angles due to the occlusion of branches or other leaves, this study conducted horizontal mirror inversion and rotation operations on the training samples. The rotated image is intercepted in the center. After rotating, the object near the edge in the image will be discarded if it is incomplete or completely lost.

Dataset Preparation

The process of dataset preparation is shown in **Figure 3**.

Datasets and sample size are shown is **Table 2**. In datasets A and B, the number of bounding boxes for tomato diseases and pests was 21,038, and the target foreground area (2,987) was marked for occlusion and overlapping tomato diseases and pests leaves. In dataset C, the number of annotations for the bounding box of tomato diseases and pests was 1,73,304, and the number of annotations for the foreground area was 24,158.

Considering that the test set is used to evaluate the detection accuracy of the model, the original image annotated by the bounding box is used as the test set.

METHOD OF IMPROVING YOLOv3-TINY

Principle of YOLOv3-Tiny

YOLO detection (Redmon et al., 2015) has developed three generations, and many networks for specific scenes have been derived. YOLO first uses the idea of regression to classify image objects, and the detection speed reaches 45 frames/s. The disadvantage is that the detection accuracy of small objects is not high. YOLOv2 (Redmon and Farhadi, 2017) have optimized the model structure of YOLO and improved the detection speed, but the detection accuracy was not improved. YOLOv3 (Redmon and Farhadi, 2018) uses deep residual network to extract image features, as the minimum feature map for feature extraction is too large, the detection speed is reduced and the detection effect for medium or large size objects is not good.

YOLOv3-tiny (Redmon, 2018) compresses the original network version without residual layer, and only two YOLO output layers with different scales are used, which improves the detection speed and accuracy of small object detection. Since tomato diseases and pests image objects are mostly small objects, and the detection speed requirements are high, it is suitable for the basic network of this detection. It uses end-to-end object detection, while ensuring accuracy, and it can greatly improve the detection speed.

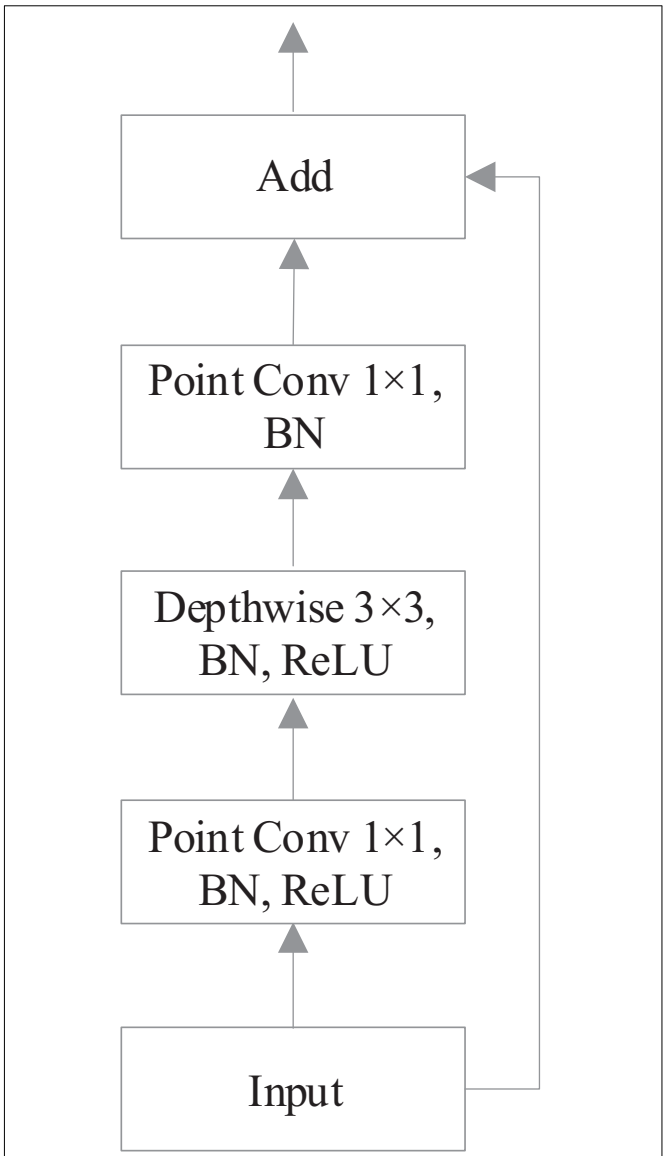


FIGURE 4 | The structure of the inverse-residual block.

TABLE 3 | Inverse-residual block parameters.

Input	Operation	Output
$h \times w \times k$	1 × 1 point conv, ReLU	$h \times w \times 2k$
$h \times w \times 2k$	3 × 3/s depthwise conv, ReLU	$h/s \times w/s \times 2k$
$h/s \times w/s \times 2k$	1 × 1 point conv, Linear	$h/s \times w/s \times 2k$

Existing Problems of YOLOv3-Tiny

In the feature extraction process of YOLOv3-Tiny model, the number of network layers in the backbone network is small, the extracted feature information is less effective, and the extraction effect is poor. Therefore, each region in the extracted feature map should be given different weights to better perform classification task. In addition, the original model cannot make full use of the

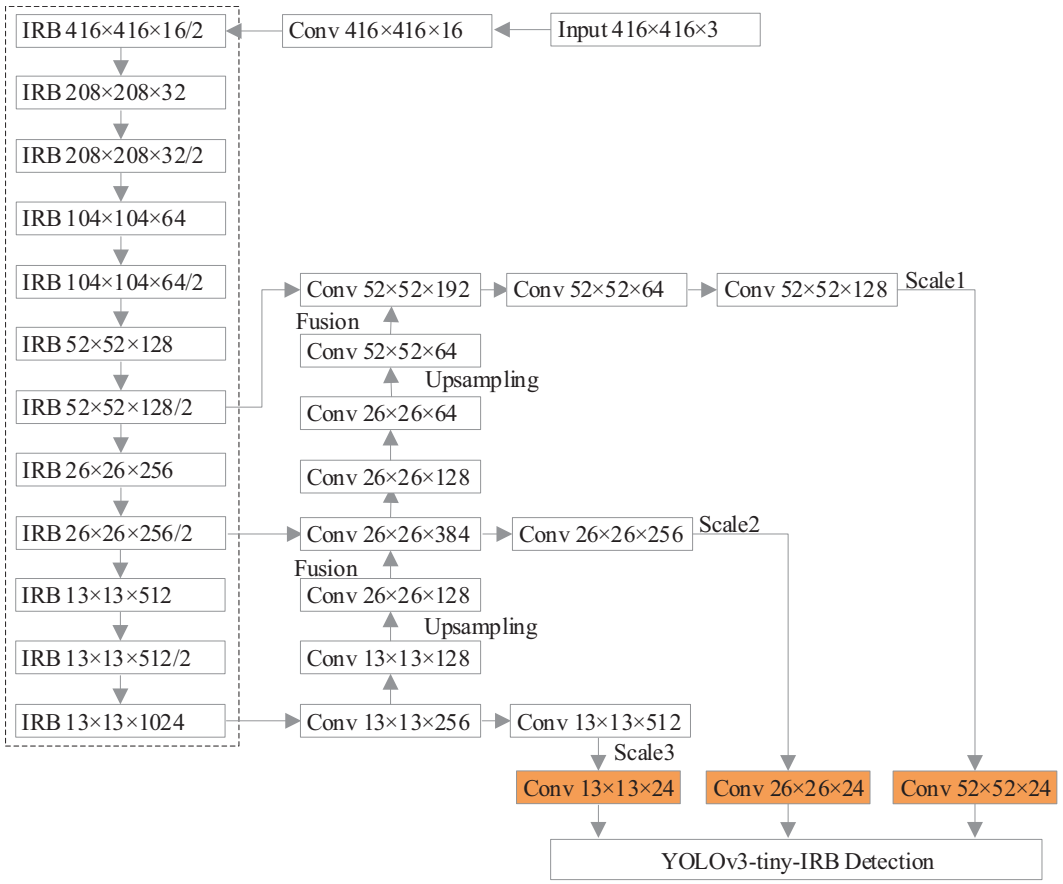


FIGURE 5 | The improved YOLOv3-tiny network model (YOLOv3-tiny-IRB).

TABLE 4 | Size and computation amount of different network models.

Network models	Model size/M	Floating point calculation amount/GFLOPs
YOLOv3	246.5	65.7
YOLOv3-tiny	34.7	5.56
YOLOv3-tiny-IRB	35.2	5.80

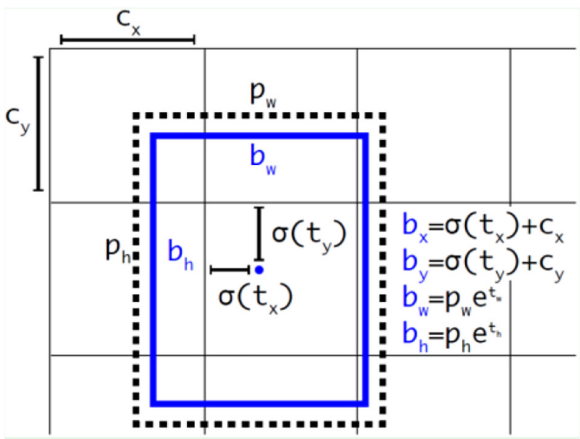


FIGURE 6 | Bounding box prediction.

feature information output from the shallow layer of the network, resulting in poor fine-grained detection ability of the model.

The Improved YOLOv3-Tiny Network (YOLOv3-Tiny-IRB)

In view of the above problems, this work improves the original network and optimizes YOLOv3-Tiny to make it more suitable for tomato diseases and pests object detection task based on field images with multiscale occlusion. In order to solve the problem that the storage and computation of conventional convolution parameters multiply with the deepening of network layers, resulting in the increase of model size and difficult application in hardware platforms with limited computing resources, this work introduces the idea of residual blocks in Resnet (He et al., 2016). Instead of conventional convolution, depth-wise separable convolution is applied to construct inverse-residual block, which transforms the "spatial cross-channel" features learning process into two parts: spatial feature learning and channel combination. Specifically, one is that depth-wise separable convolution performs spatial convolution independently on each

TABLE 5 | Experimental hardware environment configuration.

Hardware name	Model	Number
Main board	ASUS WS X299 SAGE	1
CPU	INTEL I7-9800X	1
Memory	Kingston 16G DDR4	2
Graphics card	GEFORCE GTX1080Ti	2
Solid-state hard disk	Kingston 256G	1
Hard Disk	Western Number 1T	1

input channel; the other is that point where convolution maps the output results of depth-wise separable convolution to a new channel space. The structure of the inverse-residual block is shown in **Figure 4**.

According to **Figure 4**, firstly, the input step size Stride = 1, the number of channels is adjusted by 1 × 1 convolution kernel, and the results are obtained by batch normalization (BN) algorithm and rectified linear unit (ReLU) activation function in turn; secondly, the network features are extracted by 3 × 3 convolution kernel, and pass through the ReLU function of BN algorithm; thirdly, the number of channels is adjusted by 1 × 1 convolution kernel to get the output through BN algorithm. Finally the output is added to the input before entering the structure. The structure of inverse-residual block is different from the residual in ResNet. ResNet first reduces dimension, then convolutes, and finally increases dimension, whereas inverse-residual block first increases dimension, then convolutes, and finally reduces dimension. The 1 × 1 convolution dimension enhancement is used to increase the expressive ability of the model. When the channel information is processed with the ReLU function, the channel will inevitably lose information. When there are enough channels, the lost information of one channel may still remain in other channels, so it is necessary to increase the dimension of the features first. The input of the inverse-residual block structure already contains all of the necessary information, so the ReLU activation layer is not added after the final 1 × 1 convolution to prevent information loss. After the dimension is increased, the information is more abundant. At this time, the ReLU function is added to increase the sparsity of the network. After dimension reduction, the necessary information can be maintained without loss.

The calculation of inverse-residual block in this study is shown in **Table 3**.

In **Table 3**, h and w are the height and width of the feature map, respectively, k is the number of channels of the feature map, t is the multiple of the number of expanded channels, and s is the step size. According to **Table 2**, both point convolution and depth-wise convolution of the extended channel in the inverse-residual structure of this study apply ReLU non-linear activation function. When the point convolution layer for the number of combined channels uses ReLU activation function, the negative values will be changed to 0, thus losing part of the information, and the linear activation function is used to solve the information loss problem in the process of combined channels.

The improved YOLOv3-tiny network is denoted as YOLOv3-tiny-IRB, and the network structure is shown in **Figure 5**, where the IRB (Sandler et al., 2018) is an Inverse Residual Block and the dotted box is the part of network feature extraction.

According to **Figure 5**, in the feature extraction network, the feature extraction quantity is improved by increasing the convolution layer, and the convolution with step size of 2 is used to replace the maximum pooling layer in the original network for downsampling. The inverse residual block constructed by depth-wise separable convolution is used instead of traditional convolution. The improved network is composed of 12 inverse-residual blocks, which extract high-dimensional features through inverse-residual blocks, expand feature map channels, and then carry out channel dimension reduction to obtain feature maps to make up for the deficiency of the algorithm in occlusion object detection and improve the accuracy of the algorithm. While increasing feature extraction, the model size and parameter calculation amount are effectively reduced. At the same time, there is downward transmission among scales, and the scale diversity caused by different degree of occlusion and depth of visual field decides to add an upper sampling layer on the basis of the two-scale prediction objects of the original network, which forms a three-scale prediction of 52 × 52, 26 × 26, 13 × 13. Fusion of different size features is conducive to the different object sizes in occlusion scenarios, preventing overfitting and further improving the accuracy of object detection.

Table 4 lists the size and computation amount required to process an image of YOLOv3, YOLOv3-tiny, and the network model improved in this work. It can be seen that the network model improved in this study is only 0.5M larger than YOLOv3-tiny, and the amount of computation required to process an image increases by 0.24GFLOPs, which is much smaller than that of YOLOv3 model. It has great advantages in model size and calculation amount. It meets the real-time detection requirements of the embedded system.

Anchor Parameter Optimization

When studying object detection, appropriate anchor value can improve the detection accuracy and speed. The anchor value in the original YOLO algorithm is calculated by K-means clustering method, which is more accurate than manual calculation. However, for the dataset of tomato diseases and pests in this work, the anchor value obtained by the original algorithm using COCO and VOC datasets with too large instance size is too large, so it is necessary to recalculate the appropriate anchor value according to the actual data. In tomato diseases and pests detection, clustering is to maximize the IOU value of the ratio of anchor box to ground truth, so IOU is used as the objective function to determine the distance, and its formula is as follows:

$$d(box, centroid) = 1 - IOU_{centroid}^{box} \quad (1)$$

Therefore, we set the center of centroid as the cluster center in each instance label, and BOX as the bounding box. The smaller the IOU, the larger the distance.

According to the label information of all the examples in the study, new anchor values are obtained, which are: (10, 12), (22, 24), (30, 32), (69, 86), (83, 105), (119, 192), (168, 264), (223, 296),

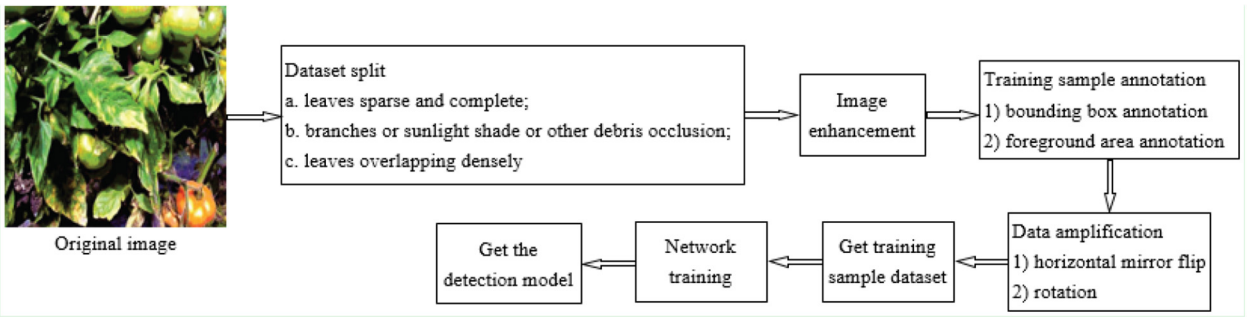


FIGURE 7 | Training flow of tomato diseases and pests object detection network.

TABLE 6 | The pseudocode of training YOLOv3-tiny-IRB.

Input: Training data

Initialize: W,b
for each batch sample X do
for $l \leftarrow 1$ to L do
if l = *Convolutional Layer* then
$\tilde{Z}^l \leftarrow BN(Z^l)$
$A^l \leftarrow \sigma\left(\tilde{Z}^l + b\right)$
else if l = *Pooling Layer* then
$A^l \leftarrow pool(A^{1-1})$
end if
end for
$J \leftarrow \frac{1}{m}\sum_{i=1}^{m} L(y^i, \tilde{y}^i)$
for $l \leftarrow L$ to 1 do
$\theta^l \leftarrow \{w^l, b^l, \gamma^l, \beta^l\}$
$V_{d\theta}^l \leftarrow \lambda \cdot V_{d\theta}^l + (1-\lambda)\frac{\partial J}{\partial \theta^l}$
$\theta^l \leftarrow \theta^l - \alpha \cdot V_{d\theta}^l$
end for
end for
Return W, b, γ, β

(311, 358). Three groups of smaller anchor boxes are assigned to larger size feature maps for predictive use; three groups of middle size anchor boxes are assigned to medium size feature maps for predictive use. In addition, three groups of anchor boxes with larger area are allocated to smaller size feature map prediction.

Each grid uses the method of directly predicting relative position to calculate three prediction boxes, as shown in **Figure 6**.

The relevant formulas in Figure 6 are as follows:

$$b_x = \sigma(t_x) + c_x \tag{2}$$

$$b_y = \sigma(t_y) + c_y \tag{3}$$

$$b_w = p_w e^{t_w} \tag{4}$$

$$b_h = p_h e^{t_h} \tag{5}$$

In the above-mentioned formulas, c_x and c_y represents the upper-left coordinates of each grid. Here, p_w and p_h represent the width and height of mapping from the anchor to the feature map, respectively, and t_x, t_y, t_w, t_h are the goals of model learning.

NETWORK TRAINING

Experimental Running Environment

The experimental hardware environment of this study is shown in **Table 5**. On this basis, the software environment is built: Ubuntu 16.04, Python, OPPENCV, CUDA, etc. The framework uses Caffe and darknet-53 framework.

Model Training Process

The training process of tomato diseases and pests object detection network is shown in **Figure 7**. After the original image in the training set is equalized by adaptive histogram, the training samples are manually annotated, including bounding box annotation and foreground area annotation; the annotated samples are expanded; and the multiscale training strategy is used for training.

The pseudocode of training YOLOv3-tiny-IRB is shown in **Table 6**. The number of network layers is L, the weight of the network is W,b. BN represents batch normalization operation. Also γ and β are the parameters, and they should be updated iteratively in the back propagation process, and λ and α represent momentum value and learning rate, respectively.

In the training phase, an asynchronous random gradient with a momentum term of 0.9 was used, the initial learning rate of the weights was 0.001, and the attenuation coefficient was set to 0.0005. In view of the differences in object scales of tomato diseases and pests in complex natural scenarios, and since individual object scales are of small size, the network training mainly adopts two strategies. One is to increase the input scale and fine-tune the network at 512 × 512 resolution to adapt to higher input resolution in detection. This strategy can improve the detection accuracy, but also reduce the detection speed. The second strategy is multiscale training. In the training iteration, the network runs every 10 batches from the set multiscale set {384, 416...672} and continue training by replacing one scale randomly again. This strategy makes the model have better detection effect at different input resolution to adapt to multiscale object detection of tomato diseases and pests. The loss descent

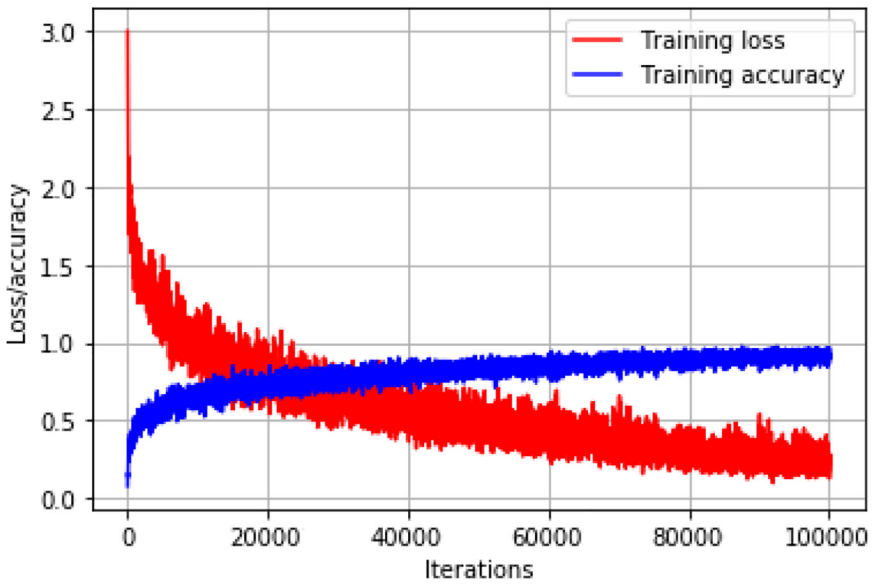


FIGURE 8 | The loss and accuracy curve during training process.

TABLE 7 | Detection results of different algorithms.

Model name	mAP (%)	F1 score	Detection speed (FPS)
DPM	73.2	0.792	0.3
Faster R-CNN	86.6	0.881	4
Mask R-CNN	87.1	0.889	3.6
SSD	85.3	0.862	55
YOLOv3	88.8	0.897	62
YOLOv3-tiny	88.1	0.893	220
YOLOv3-tiny-IRB	**93.1**	**0.922**	**206**

TABLE 8 | Detection results on different training sets.

Training set	mAP (%)	F1 score
A	90.3	0.901
B	92.6	0.913
C	**93.1**	**0.922**

curve versus the prediction accuracy curve of training set during training process is shown in **Figure 8**.

According to **Figure 8**, the accuracy curve is rising steadily whereas the loss curve is decreasing. The accuracy curve gradually, leveled off after 80,000 iterations, and the model at 1,00,000 iterations were selected for this study.

EXPERIMENTAL RESULTS AND COMPARATIVE ANALYSIS

The object detection network of tomato pests and diseases was trained with training sets A, B, and C (see **Table 1**), respectively, and the performance of object detection of tomato diseases and pests in the scene of occlusion and leaf overlap was analyzed, and

TABLE 9 | Detection results by training set with foreground region.

Group	Sample numbers	Annotation numbers		mAP (%)	F1 score
		Bounding box	Foreground region		
a	3500	21038	0	88.6	0.899
b	3500	18051	2987	91.7	0.908
c	29016	173304	0	92.8	0.912
d	**29016**	**149116**	**24158**	**94.2**	**0.936**

compared with the performance of the detection models such as Faster RCNN (Ren et al., 2017), YOLO and Adaboost. Around 1,500 images (9,067 tomato diseases and pests objects) were input into the trained network for location regression. When the IOU (intersection and convergence ratio) of the object bounding box predicted by the model and the manually labeled bounding box is more than 0.7, the detection is considered correct, otherwise it is wrong and the test results are obtained.

Precision (P), recall (R), F1 value, and detection speed were selected as evaluation criteria. Sample S is divided into four types according to the combination of the true category of sample S and the predicted category of model: True positive (TP) represents the number of correctly classified positive samples, FP represents the number of incorrectly classified positive samples, false negative (FN) represents the number of incorrectly classified negative samples (FN), and true negative (TN) represents the number of correctly classified negative samples.

Precision (P) represents the proportion of samples that are truly positive in all samples that are predicted to be positive, and the formula is

$$P = \frac{TP}{TP + FP} \tag{6}$$

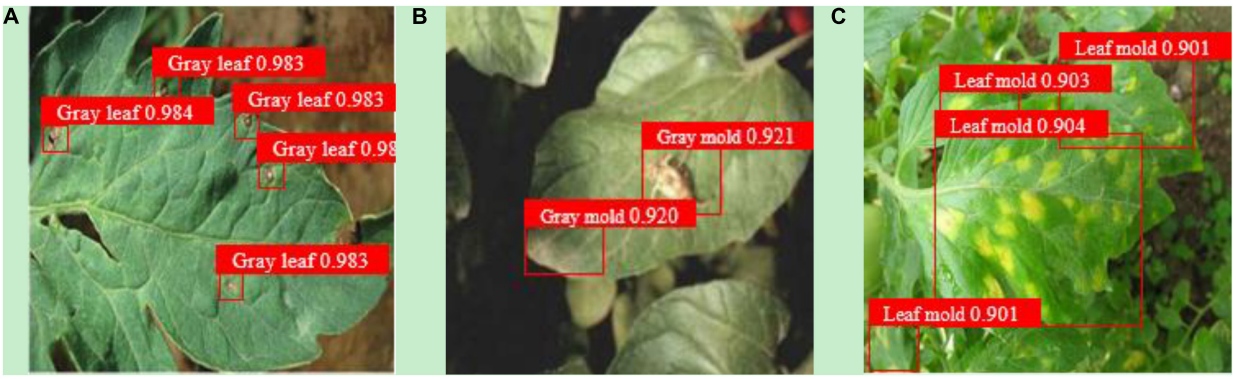


FIGURE 9 | The detection effect diagram of YOLOv3-tiny-IRB [**(A)** deep separation; **(B)** debris occlusion; **(C)** leaves overlapping].

Recall ® represents the proportion of samples that are predicted to be positive of the truly positive samples. The formula is

$$R = \frac{FP}{FP + TN} \tag{7}$$

The F1 value is a measure function of balancing precision P and recall R, and the calculation formula is

$$F1 = \frac{2PR}{P + R} \tag{8}$$

In object detection, in each category the P–R curve can be drawn according to precision P and recall R. The average accuracy AP value of single category detection is the area between P–R curve and coordinate axis, and the calculation formula is as follows:

$$AP = \int_0^1 P(R)dR \tag{9}$$

The average of AP values of all categories is mAP, and the formula is

$$mAP = \frac{1}{c}\sum AP \tag{10}$$

In the above-mentioned formula C is the number of categories contained in the dataset.

Frame rate per second (FPS) is a common indicator of speed, which is the number of images that can be processed per second.

Performance Comparison of Several Different Algorithms

The dataset after data processing was used as training set. Deformable Parts Model (DPM) (Sun et al., 2014), Faster R-CNN, Mask R-CNN, single shot multibox detector (SSD) (Liu et al., 2016), YOLOv3, YOLOv3-tiny and YOLOv3-tiny-IRB are taken as a basic network for training and testing respectively.

The test results of different algorithms on the test set are shown in **Table 7**.

It can be seen that the detection accuracy of YOLOv3-tiny-IRB in this work is much higher than other models. The accuracies of Faster R-CNN, Mask R-CNN, SSD, YOLOv3, YOLOv3-tiny, and YOLOv3-tiny-IRB, which use CNN for convolution feature extraction, are significantly higher than that of DPM algorithm using HOG feature. Traditional object detection algorithm relies on manual designed features, which uses sliding window to select candidate boxes, resulting in severe window redundancy problem and poor generalization performance of feature extraction methods. As a result, the detection accuracy is low and the algorithm steps are numerous, which leads to the slow detection speed and poor real-time performance. Since CNN can simultaneously extract color, texture and shape features, it is superior to traditional methods, and so the performance of CNN detection method is superior.

According to **Table 6**, the detection speed of DPM detection method in the traditional mainstream machine learning algorithm is the slowest. Faster R-CNN and Mask R-CNN algorithm generates more than 2,000 object candidate region by region, generating a network in the detection process, and then classifies candidate regions by CNN, whereas YOLO series algorithms directly process the whole image by CNN, which reduces the computational complexity, so the detection speed is faster than Faster R-CNN and Mask R-CNN. YOLOv3-tiny is faster than YOLOv3 detection, but there are only two levels of detection and no fusion of small objects, so there is no way to identify objects of different scales well.

Compared with the original YOLOv3 and YOLOv3-tiny, the mAP improved by 4.3 and 5.0%, respectively. The introduction of the inverse-residual module improved the ability of network to extract features and increased the participation of finer feature maps in location regression and classification, thus facilitating the improvement of the detection accuracy of YOLOv3-tiny-IRB. Meanwhile, the inverse-residual block had little effect on the detection speed, and the speed reached 206 frames/s. Therefore, it maintains a good real-time performance while improving the detection accuracy. Overall, YOLOv3-tiny-IRB can achieve trade-off of accuracy and speed, so that the model can be deployed on a large scale in hardware platforms such as embedded devices and mobile terminals to meet the actual needs.

Detection Results of Amplified Datasets

Based on YOLOv3-tiny-IRB network, the comparison test before and after data amplification was carried out. As can be

TABLE 10 | Detection result comparison.

Object detection scenarios	mAP (%)	F1
(a) Deep separation	98.3	0.971
(b) Debris occlusion	92.1	0.915
(c) Leaves overlapping	90.2	0.901

seen from **Table 8**, compared with the original image dataset (Training set A), the mAP and F1 score of the model of the enhanced dataset (Training set B) were increased by 2.3 and 0.012%, respectively. After data amplification (Training set C), The mAP and F1 score of the model were improved by 2.8 and 0.021%, respectively compared with the preamplification (Training set A). The results showed that image enhancement, mirror, rotation, and other processing methods could further improve the detection accuracy.

Detection Results Using Object Foreground Training Samples

To verify the effect of foreground region training samples on detection accuracy, a comparative experiment was conducted. The 3,500 training samples in the original image were divided into two groups: group A training set contained only 21,038 bounding box annotations for all samples, group B contained 2,987 foreground region annotations for occluding and overlapping samples, and 18,051 bounding box annotations for uncovered samples. The training set of group C contains 1,73,304 bounding box annotations, and group D contains 24,158 foreground region annotations, and 1,49,146 bounding box annotations. YOLOv3-tiny-IRB is trained with these four training sets, and the test results on the test set are shown in **Table 9**. It can be seen that the object foreground region of the tomato diseases and pests training sample is annotated with YOLOv3-tiny-IRB network training. The detection accuracy is obviously improved and the difficulty of occlusion and overlap detection is overcome by reducing the interference of the features of the non-foreground region in the boundary frame. After data amplification, the detection accuracy of the model obtained by using the training set marked by the foreground region is significantly improved compared with the model without foreground region labeling, and the mAP and F1 score of all objects in the test set are improved by 1.4% and 0.024, respectively.

Detection Results Under Conditions of Occlusion and Overlapping

Under the different object detection scenarios of (a) deep separation, (b) debris occlusion, and (c) leaves overlapping, YOLOv3-tiny-IRB trained with dataset C can achieve good detection performance, as shown in **Figure 9** and **Table 10**. It can be seen that the network model designed in this work can detect tomato diseases and pests under a certain degree of occlusion interference and dense overlap of leaves. For the detection of leaves overlapping scenario, YOLOv3-tiny-IRB still reaches 90.2% of mAP, but its detection accuracy is significantly lower than that under deep separation and debris occlusion scenarios.

TABLE 11 | Detection results of each species of diseases and pests.

Species	Precision (%)	Recall (%)	F1 score
Early blight	93.9	86.5	0.922
Late blight	92.4	85.8	0.901
Gray leaf spot	93.5	86.4	0.912
Brown spot	92.7	84.2	0.910
Coal pollution	93.9	86.1	0.926
Gray mold	94.5	86.9	0.928
Leaf mold	94.8	87.1	0.925
Powdery mildew	92.8	84.3	0.917
Leaf curl	93.2	87.2	0.919
Mosaic	91.1	82.6	0.920
Leaf miner	90.2	82.7	0.904
Greenhouse whitefly	90.1	83.3	0.903
Total	93.9	86.5	0.922

Detection Results of Each Species of Diseases and Pests

To discuss detection results of each species of diseases and pests, this study compared the performance of 12 tomato diseases and pests using the improved model. In the original data, there are certain similar symptoms of the disease with similar colors. Pests are easy to discern but are densely distributed, making it more difficult to be fully detected. The detection results of each species of diseases and pests are shown in **Table 11**.

The results showed that the improved model performed excellent in detection accuracy. Detection of twelve different types of diseases and pests all achieved good results, with all F1 scores reaching more than 0.9, and the detection time also reached the requirement of real-time performance. Therefore, the improved model has good generalization ability and can adapt to the needs of rapid detection of tomato pest and disease under natural environmental conditions.

CONCLUSION AND FUTURE DIRECTIONS

The experimental results show that the proposed YOLOv3-tiny-IRB algorithm takes into account the simultaneous improvement of detection accuracy and speed, and improves the real-time detection of multiscale objects of occlusion or overlapping tomato diseases and pests in complex natural environment. The research of real-time detection algorithm in complex scenarios can better serve the needs of early warning of plant diseases and pests in smart agriculture. This work not only improves the performance of YOLOv3-tiny network in occlusion or overlapping tomato diseases and pests, but also provides a new method for other object detection, such as fruit harvesting robot, field rabbit, and field bird recognition.

At present, there are many kinds of plant diseases and pests. How to identify more kinds of plant diseases and pests through feature extraction and network structure adjustment and improve the accuracy and efficiency of identification is the direction of follow-up research.

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/s.

AUTHOR CONTRIBUTIONS

XW designed the research. JL and XW conducted the experiments and data analysis and wrote the manuscript. GL and XW revised the manuscript. All the authors read and approved the manuscript.

FUNDING

This study was supported by the Facility Horticulture Laboratory of Universities in Shandong (2019YY003), Key Research and Development Plan of Shandong Province (2020RKA07036 and 2019GNC106034), Shandong social science planning project (21CPYJ20), Natural Science Foundation of Shandong Province (ZR2021QC173), and Weifang Science and Technology Development Plan (2021GX054).

ACKNOWLEDGMENTS

Appreciations are given to the editor and reviewers of the Journal.

REFERENCES

Algorithms, L. O. (2015). "Adaptive histogram equalization," in *Proceedings of the International Conference on Electrical and Electronics Engineering, ELECO 2009* (Piscataway, NJ: IEEE).

Anagnostis, A., Asiminari, G., Papageorgiou, E., and Bochtis, D. (2020). A convolutional neural networks based method for anthracnose infected walnut tree leaves identification. *Appl. Sci.* 10:469. doi: 10.3390/app10020469

Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. doi: 10.3389/fpls.2020.599886

Cai, Z., and Vasconcelos, N. (2017). Cascade R-CNN:delving into high quality object detection. *arXiv* [Preprint]. arXiv:1712.00726,

Dai, J., Li, Y., He, K., and Sun, J. (2016). "R-fcn:object detection via region based fully convolutional networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, (Red Hook, NY: Curran Associates Inc), 379–387.

Ding, W., and Taylor, G. (2016). Automatic moth detection from trap images for pest management. *Comput. Electron. Agric.* 123, 17–28. doi: 10.1002/ps.6433

Fuentes, A., Yoon, S., and Park, D. S. (2019). Deep learning-based phenotyping system with glocal description of plant anomalies and symptoms. *Front. Plant Sci.* 10:1321. doi: 10.3389/fpls.2019.01321

Geetharamani, G., and Arun Pandian, J. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Comput. Electric. Eng.* 76, 323–338.

Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, (Washington, DC: IEEE), 1440–1448.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2013). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, (Columbus, OH: IEEE).

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the 2017 IEEE International Conference on Computer Vision(ICCV)*, (Venice: IEEE), 2980–2988.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV: IEEE).

Li, J. (2012). Research on tomato bacterial pith necrosis. *Plant Dis. Pests* 3, 9–11.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. *arXiv* [Preprint]. arXiv:1612.03144,

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2018). Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 318–327. doi: 10.1109/tpami.2018.2858826

Liu, G., Nouaze, J. C., Touko, P. L., and Kim, J. H. (2020). Yolo-tomato: a robust algorithm for tomato detection based on yolov3. *Sensors* 20:2145. doi: 10.3390/s20072145

Liu, J., and Wang, X. (2020a). Early recognition of tomato gray leaf spot disease based on mobilenetv2-yolov3 model. *Plant Methods* 16:83.

Liu, J., and Wang, X. (2020b). Tomato diseases and pests detection based on improved yolo v3 convolutional neural network. *Front. Plant Sci.* 11:898. doi: 10.3389/fpls.2020.00898

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., et al. (2016). Ssd: single shot multibox detector . *arXiv* [Preprint]. arXiv:1512.02325,

Ouhami, M., Es-Saady, Y., Hajji, M. E., Hafiane, A., and Yassa, M. E. (2020). "Deep transfer learning models for tomato disease detection," in *Image and Signal Processing*, eds A. El Moataz, D. Mammass, A. Mansouri, and F. Nouboud (Cham: Springer).

Prabhakar, M., Purushothaman, R., and Awasthi, D. P. (2020). Deep learning based assessment of disease severity for early blight in tomato crop. *Multimed. Tools Appl.* 79, 28773–28784. doi: 10.1007/s11042-020-09461-w

Redmon, J. (2018). *YOLO-tiny. [CP/DK]*. Available online at: https://github.com/pjreddie/darknet/blob/master/cfg/yolov3-tiny.cfg (accessed April 25, 2021).

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2015). You only look once: unified, real-time object detection. *arXiv* [Preprint] arXiv:1506.02640,

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, NV: IEEE), 779–788.

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the 2017 IEEE Conference on Computer Vision & Pattern Recognition*, (Honolulu, HI: IEEE), 6517–6525.

Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. *arXiv* [Preprint] arXiv arXiv:1804.02767,

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.* 2017, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Salt Lake City, UT: IEEE).

Shekhawat, R. S., and Sinha, A. (2020). Review of image processing approaches for detecting plant diseases. *IET Image Process.* 14, 1427–1439. doi: 10.1049/iet-ipr.2018.6210

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

Sun, C., Wang, X., and Lu, P. (2014). "Object ranking on deformable part models with bagged LambdaMART," in *Computer Vision – ACCV 2014*, eds D. Cremers, I. Reid, H. Saito, and M. H. Yang (Cham: Springer International Publishing).

TzuTa, L. (2017). *LabelImg [CP/DK]*. Available online at: https://github.com/tzutalin/labelImg (accessed April 25, 2021).

Wang, X. Y., Feng, J., Yang, Y. Y., and Zhu, X. P. (2018). Natural occurrence of tomato chlorosis virus in cowpea (vigna unguiculata) in china. *Plant Dis.* 102:254. doi: 10.1094/pdis-02-17-0288-pdn

Xu, Z. F., Jia, R. S., Liu, Y. B., Zhao, C. Y., and Sun, H. M. (2020). Fast method of detecting tomatoes in a complex scene for picking robots. *IEEE Access* 8, 55289–55299.

Zhang, S., Wen, L., Bian, X., Lei, Z., and Li, S. Z. (2017). Single-shot refinement neural network for object detection. *arXiv* [Preprint] arXiv:1711.06897,

Zhao, J., and Qu, J. (2019a). "Healthy and diseased tomatoes detection based on YOLOv2," in *Human Centered Computing. HCC 2018. Lecture Notes in Computer Science*, Vol. 11354, eds Y. Tang, Q. Zu, and J. Rodríguez García (Cham: Springer), doi: 10.1007/978-3-030-15127-0_34

Zhao, J., and Qu, J. (2019b). "A detection method for tomato fruit common physiological diseases based on YOLOv2," in *Proceedings of the 2019 10th International Conference on Information Technology in Medicine and Education (ITME)*, (Qingdao: IEEE).

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 December 2021
doi: 10.3389/fpls.2021.789911

Multi-Threshold Image Segmentation of Maize Diseases Based on Elite Comprehensive Particle Swarm Optimization and Otsu

Chengcheng Chen[1,2], *Xianchang Wang*[1,2,3*†], *Ali Asghar Heidari*[4], *Helong Yu*[5*†] and *Huiling Chen*[6*†]

[1] College of Computer Science and Technology, Jilin University, Changchun, China, [2] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Changchun, China, [3] Chengdu Kestrel Artificial Intelligence Institute, Chengdu, China, [4] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, [5] College of Information Technology, Jilin Agricultural University, Changchun, China, [6] College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, China

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Hossein Moayedi,
Southern Illinois University Edwardsville, United States
Hongbin Pei,
Xi'an Jiaotong University, China

***Correspondence:**
Xianchang Wang
xcwang89@jlu.edu.cn
Helong Yu
yuhelong@aliyun.com
Huiling Chen
chenhuiling.jlu@gmail.com

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *11 October 2021*
Accepted: *01 November 2021*
Published: *13 December 2021*

Citation:
Chen C, Wang X, Heidari AA, Yu H and Chen H (2021) Multi-Threshold Image Segmentation of Maize Diseases Based on Elite Comprehensive Particle Swarm Optimization and Otsu. Front. Plant Sci. 12:789911. doi: 10.3389/fpls.2021.789911

Maize is a major global food crop and as one of the most productive grain crops, it can be eaten; it is also a good feed for the development of animal husbandry and essential raw material for light industry, chemical industry, medicine, and health. Diseases are the main factor limiting the high and stable yield of maize. Scientific and practical identification is a vital link to reduce the damage of diseases and accurate segmentation of disease spots is one of the fundamental techniques for disease identification. However, one single method cannot achieve a good segmentation effect to meet the diversity and complexity of disease spots. In order to solve the shortcomings of noise interference and oversegmentation in the Otsu segmentation method, a non-local mean filtered two-dimensional histogram was used to remove the noise in disease images and a new elite strategy improved comprehensive particle swarm optimization (PSO) method was used to find the optimal segmentation threshold of the objective function in this study. The experimental results of segmenting three kinds of maize foliar disease images show that the segmentation effect of this method is better than other similar algorithms and it has better convergence and stability.

Keywords: non-local mean filtering, enhanced comprehensive learning particle optimizer, Otsu, multi-threshold image segmentation, maize disease image

HIGHLIGHTS

- The Otsu is combined with an elite comprehensive particle swarm algorithm for image segmentation.
- A non-local mean filtered 2D histogram is combined to remove noise.
- The GCLPSO is the first time applied to the segmentation of a variety of diseases of maize.
- The GCLPSO performs better than other similar algorithms under multi-threshold segmentation.

INTRODUCTION

Diseases often occur during maize cultivation and failure to prevent and control them in time can decrease corn yield and quality, making a loss of economic benefits for the grower. According to the data reported by the Food and Agriculture Organization (FAO) of the United Nations, the annual natural loss rate caused by agricultural pests and diseases is more than 37% and agricultural pest and disease identification and control are of great importance to improve agricultural production (Dhami et al., 2015). Traditional manual identification methods rely on experience, high labor costs, subjective factors, and lack of accuracy (Gao and Lin, 2019). In recent years, computer vision and image processing methods have gradually developed; the method is more objective and supports the real-time online diagnosis, bringing new opportunities to solve agricultural disease diagnosis, reduce economic losses, and improve crop yields (Wang et al., 2019).

Diseases are the main causal factors affecting crop growth and in order to improve the quality and yield of crops, targeted screening and diagnosis are needed during the growth process. Crop disease recognition technology based on machine vision image processing usually includes disease image preprocessing, image segmentation, feature extraction, image recognition, and interpretation (Iqbal et al., 2018). The higher the accuracy of image segmentation, the higher the accuracy of feature extraction and image recognition. Therefore, higher accuracy segmentation methods are the main research direction for scholars in disease identification and diagnosis. The mainstream agricultural disease image segmentation techniques are mainly based on clustering methods (Wang et al., 2018), region growth methods (Jothiaruna and Sundar, 2020), edge detection methods (Shaikh et al., 2017), threshold segmentation methods (Wang et al., 2013), heuristic algorithm methods (Zhou et al., 2018), machine learning, and deep learning methods (Ale et al., 2019). Although there are many novel and effective image segmentation methods in recent literature (Elaziz et al., 2020; Rodriguez-Esparza et al., 2020; Zhao et al., 2020a,b, 2021), there is not yet a general robustness well-adapted segmentation method that can be applied to multiple scenes due to the interference of many crop species, complex background information, diverse and disorderly disease spot morphological texture colors, multiple disease spot interference, blurred leaf surface texture, and disease spot boundaries. The currently available segmentation methods and their advantages and disadvantages are shown in following **Table 1**.

Threshold-based segmentation methods are computationally efficient, straightforward, and widely used in multiple fields and crop image recognition. Subramani et al. (2019) presented a method that combines non-local median filter and double line clustering to analyze the anthracnose, blight disease in grapes, tomato, and cucumber. Xiong et al. (2020) proposed an automatic image segmentation algorithm (AISA) based on the GrabCut algorithm that automatically removes the background information of the images while retaining the disease spots. Kumari et al. (2019) presented a novel approach based on the simple linear iterative clustering segmentation method to detect disease in plant leaves. Yan et al. (2018) proposed that extract the H channel information in the HSI, more common, components hue (H), saturation (S), brightness(I) and use each pixel and its local average value to form a two-dimensional (2D) histogram, then segment color space the image by the optimal threshold of the Otsu algorithm in the polar diameter information to improve the segmentation accuracy of rice blast images. Zhang et al. (2019) presented a novel hybrid clustering segmentation method of plant disease leaf image. Hu et al. (2017) proposed an improved Chan–Vese (C-V) model for wheat leaf lesion segmentation. Gao and Lin (2019) proposed a fully automatic segmentation method using leaf images of medicinal plants in complex backgrounds with vein enhancement and extraction in the images. Among the threshold image segmentation methods, the Otsu (Merzban and Elbayoumi, 2019) segmentation method is one of the classical threshold segmentation methods, which has obvious disadvantages of misclassification and computational complexity using grayscale histograms and cannot be well-adapted to complex and diverse crop disease images. Mittal and Saraswat (2018) proposed a 2D histogram with non-local mean filtering, which can effectively reduce the loss of image details. However, the Otsu method finds the optimal threshold using the exhaustive method, the complexity grows exponentially with the increase of the number of thresholds, and there are obvious shortcomings in the computational performance. The application of swarm intelligence algorithm is a bionic approach to solve optimization problems with intelligence, parallelism, and robustness, which is widely used in thresholding optimization problems. Jia et al. (2019a) used improved moth flame optimization (MFO) for multistage thresholding segmentation of color images. Jia et al. (2019b) proposed an improved multilevel optimization algorithm based on Lévy flight for multi-threshold color image segmentation. Kotte et al. (2018) proposed a fast multi-thresholding method for gray image segmentation based on a differential evolution algorithm. Bao et al. (2019) proposed a novel hybrid Harris Hawks optimization method for multilevel threshold segmentation of color images.

In this study, non-local mean filtering of 2D histogram Otsu was used to segmentation for multi-threshold image processing and enhanced comprehensive learning particle swarm optimizer with elite-based dominance scheme (GCLPSO) (Chen et al., 2020) was used to find the optimal threshold, which was used to achieve higher convergence speed and accuracy, can quickly achieve convergence to the optimal value, and improve the efficiency of image segmentation. The methods were applied to maize leaf disease images in the Plant Village[1] public database. The experimental results indicated that the method effectively improved the segmentation of the three maize disease spot images and could obtain more apparent disease spot areas. To verify the experimental validity, the used GCLPSO was compared with the original CLPSO (Liang et al., 2006) and two of improved algorithms such as Sine Cosine Algorithm and Differential Evolution (SCADE) (Nenavath and Jatoth, 2018), modified sine cosine algorithm (m_SCA) (Qu et al., 2018) and three original algorithms such as salp swarm algorithm (SSA) (Aljarah et al., 2018; Faris et al., 2018; Abbassi et al., 2019), SCA (Oliva et al., 2018; Qu et al., 2018; Kuo et al.,

[1]Available online at: https://github.com/Tmcsn/AI-Challenger-2018-CropDisease.

TABLE 1 | Existing segmentation methods and their advantages and disadvantages.

Segmentation method	Advantages	Disadvantages
Edge detection	Fast search detection and good detection of edges	Cannot get a better regional structure; The conflict between noise immunity and detection accuracy during edge detection
Region growth	Effectively overcome the disadvantage of small continuous image segmentation space existing; Better regional characteristics	Easy to cause excessive segmentation of images, complex and computationally intensive
Clustering	High sensitivity to initial settings, and sometimes it needs human decisions within initialization	No consideration of spatial information, sensitive to noise and gray scale inhomogeneity
Threshold	Direct use of the grayscale characteristics of the image, so the calculation is simple, efficient, and fast	Sensitive to noise, not obvious to grayscale differences and different target grayscale values have overlapping segmentation is not obvious, need to find a suitable threshold with other methods
Mathematical morphology	Good positioning effect, high segmentation accuracy, good anti-noise performance	High requirement for accuracy of pre-processed images; otherwise, the speed of calculation is reduced
Deep learning	Resolve noise and unevenness in images	Requires a large amount of data, very slow, complex structure, segmentation accuracy is related to the amount of data

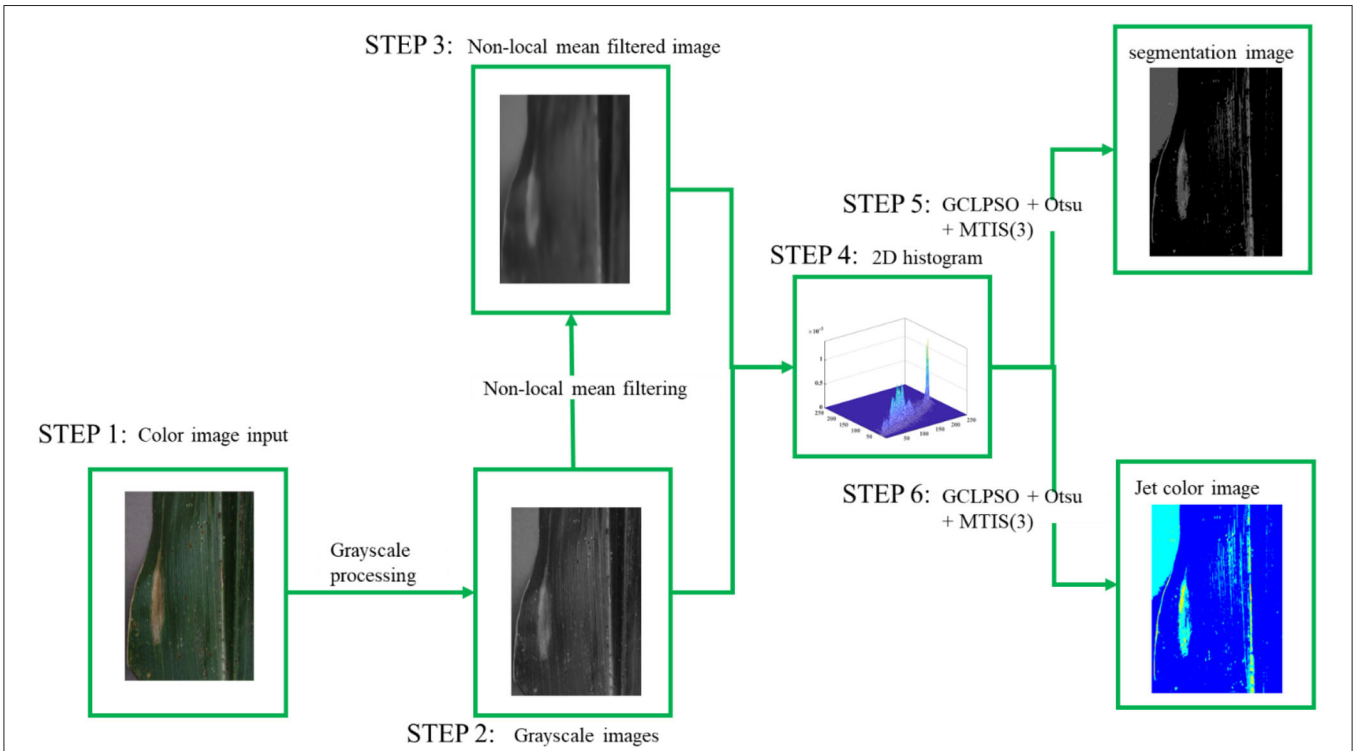


FIGURE 1 | Flowchart of multi-threshold image segmentation of maize diseases using GCLPSO + Otsu.

2020), and Slime mould algorithm (SMA) (Abbassi et al., 2019), respectively. The segmentation experiments with multiple thresholds were also performed separately. In addition, the feature similarity (FSIM) index (Zhang et al., 2011), peak signal-to-noise ratio (PSNR) (Setiadi, 2020), and structural similarity (SSIM) index (Wang et al., 2004) were used to compare the image segmentation results for evaluation and the mean, variance, and the Wilcoxon signed-rank (Garcia et al., 2010) tests were used to analyze the evaluation results. Through a series of analyses and comparisons of experimental results, GCLPSO for non-local mean filtering of 2D histogram Otsu multi-threshold image processing outperforms other algorithms in terms of overall performance and can effectively segment corn leaf disease images. **Figure 1** shows the steps of image segmentation based on this method.

The rest of this study is organized as follows. Chapter 2 introduces the multi-threshold Otsu segmentation. Chapter 3 introduces the non-local mean filtered 2D histogram. Chapter 4 introduces GCLPSO. Chapter 5 conducts a series of comparison experiments between GCLPSO and other optimal thresholding methods, and chapter 6 summarizes the whole paper and the direction of future work.

MULTI-THRESHOLD OTSU SEGMENTATION

Crop disease image segmentation is mainly concerned with separating the disease spots of a crop from the leaves or other backgrounds containing the leaves. Multi-threshold segmentation is an integral part of digital agricultural image processing. It mainly refers to marking out the targets of interest in an image by setting multiple thresholds. The selection of thresholds is critical and related to the good or bad results after segmentation. Otsu method is a more common and perfect method in multi-threshold image segmentation (MTIS), which was proposed by Japanese scholar Otsu in 1979 and it is also called the maximum interclass variance method and its principle is that the interclass variance between foreground and background images is maximum after image binarization segmentation according to the thresholds obtained by the Otsu method (Merzban and Elbayoumi, 2019).

For an $M \times N$ image I, x is the row coordinates of image pixel points, y is the column coordinates of image pixel points, where $0 \leq x < m$, $0 \leq y < n$. The image gray level is $S = \{0, 1, \ldots, L-1\}, (L = 256)$ and the number of all the image pixel points is denoted as: $m \times n$. G is the corresponding averaged image; then pixel gray level in G can be defined as follows:

$$G(x, y) = \sum_{\tilde{x}=x-\left(k-\frac{1}{2}\right)}^{x+\frac{k-1}{2}} \sum_{\tilde{y}=x-\left(k-\frac{1}{2}\right)}^{y+\frac{k-1}{2}} I(\tilde{x}, \tilde{y}) \quad (1)$$

where $I(x, y)$ and $G(x, y)$ represent the gray level of the pixel at x, y in I and G, respectively. k represents the size of the filter and the value of k is set to be 3 in this study. Let i, j be pixel gray level of original image and averaged image; then i, j is a gray level pair representing that the pixel gray level in image I is i and the gray level of the corresponding pixel at the same location in the averaged image G is j. Suppose f_{ij} is the pixel number of i, j, then the 2D probability function can be defined as:

$$P_{ij} = \frac{f_{ij}}{m \times n} \quad (2)$$

where $i,j \epsilon$ (0, L – 1) and $\sum_i \sum_j P_{ij} = 1$. The average vector of the 2D histogram is as follows:

$$\mu_T = (\mu_{Ti}, \mu_{Tj})^T = \left(\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} iP_{ij}, \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} jP_{ij} \right)^T \quad (3)$$

A given threshold pair (s, t), pixels can be partitioned into two sets, C_0 and C_1 (background and foreground) and the class occurrence probabilities can be expressed as:

$$\omega_0 = P(C_0) = \sum_{i=0}^{s} \sum_{j=0}^{t} P_{ij},$$
$$\omega_1 = P(C_1) = \sum_{i=s+1}^{L-1} \sum_{j=t+1}^{L-1} P_{ij}. \quad (4)$$

The corresponding mean vectors of C_0 and C_1 are:

$$\mu_0 = (\mu_{0i}, \mu_{0j})^T = \left(\sum_{i=0}^{s} \sum_{j=0}^{t} \frac{iP_{ij}}{\omega_0}, \sum_{i=0}^{s} \sum_{j=0}^{t} \frac{jP_{ij}}{\omega_0} \right)^T,$$
$$\mu_1 = (\mu_{1i}, \mu_{1j})^T = \left(\sum_{i=s+1}^{L-1} \sum_{j=t+1}^{L-1} \frac{iP_{ij}}{\omega_1}, \sum_{i=s+1}^{L-1} \sum_{j=t+1}^{L-1} \frac{iP_{ij}}{\omega_1} \right)^T \quad (5)$$

If we ignore the diagonal data far away from the 2D histogram in the image, we can get the following formula can be obtained:

$$\omega_0 + \omega_1 \cong 1, \ \ \mu_T = \omega_0 \mu_0 + \omega_1 \mu_1 \quad (6)$$

The expression between classes in the 2D Otsu algorithm is as follows:

$$tr(\sigma_B) = tr\left(\sum_{K=0}^{1} \omega_K [(\mu_K - \mu_T)(\mu_K - \mu_T)^T]\right) \quad (7)$$

Then, we calculate the maximum value in Equation (8) to get the best threshold pair (s*, t*):

$$(s^*, t^*) = argmax_{0<s<L, 0<t<L} [tr(\sigma_B)] \quad (8)$$

The larger the interclass variance, the closer the threshold to the correct image segmentation. The essence of threshold segmentation can be seen as an optimization problem of classifying image pixels according to multiple pixel gray levels, translating into a mathematical model problem of solving the objective function with the best quality. The Otsu thresholding segmentation uses an iterative approach to find the threshold that maximizes the between-class variance for the final desired threshold, so the complexity of algorithm grows exponentially as the number of thresholds and dimensions increases. In agricultural disease diagnosis applications, it could not meet the requirements of rapid diagnosis. Swarm intelligence optimization algorithm can quickly improve computational efficiency; many scholars at home and abroad carry out much evolutionary research.

NON-LOCAL MEAN FILTERED 2D HISTOGRAM

In reality, the occurrence of crop diseases is not a single event, but may be accompanied by damage caused by insects, multiple diseases or weather, and other disasters; in the image capture process, the different angles of shooting disease spots, sunlight, and room light cause uncertainty for segmentation; the irregular diversity of disease spots themselves is also the main noise that causes the inability to extract disease spots accurately. The noise can bring great difficulties to solve the threshold processing. A common approach that requires noise reduction at the source to enhance performance is to smooth the image in priority. Non-local means filtering (NL-means) is a novel denoising technique proposed in recent years. This method makes full use of the

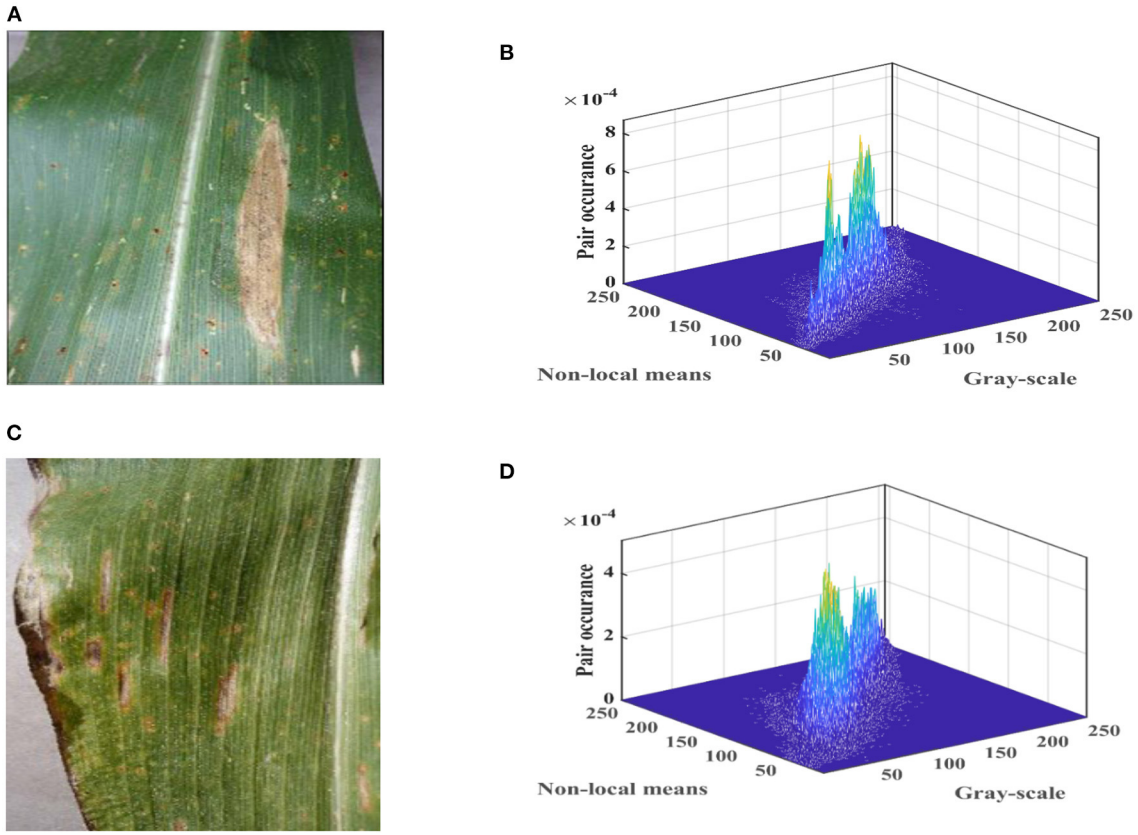


FIGURE 2 | Color images and three-dimensional (3D) views of two-dimensional (2D) histograms for maize leaf spot disease and maize gray spot disease in the Plant Village.

redundant information in the image and can maintain the maximum detail features of the image while denoising. A brief algorithm with superior performance characterizes the method. The basic idea is that the estimate of the current pixel is obtained by a weighted average of pixels in the image that has a similar neighborhood structure to it. It can effectively remove most of the noise on crop leaves for other reasons and is an effective method for removing crop disease image noise.

In image I, the grayscale values of p and q corresponding pixels are $I(p)$ and $I(q)$, respectively, then the non-local mean value of the image is calculated as follows:

$$O(p) = \frac{\sum_{q \in I} I(q)\, \omega(p,q)}{\sum_{q \in I} \omega(p,q)} \tag{9}$$

$$\omega(p,q) = \exp^{-\frac{|\mu(p)-\mu(q)|^2}{\sigma^2}} \tag{10}$$

$$\mu(p) = \frac{1}{m \times m} \sum_{i \in L(p)} I(i) \tag{11}$$

$$\mu(q) = \frac{1}{m \times m} \sum_{i \in L(q)} I(i) \tag{12}$$

$O(p)$ is the non-local mean filtered value of pixel p, $\omega(p,q)$ is the weight of p pixels and q pixels, σ is the SD, $\mu(p)$ and $\mu(q)$ are the local means of p and q, $L(p)$ is the $m \times m$ domain window around p pixels, and $L(q)$ is the $m \times m$ domain window around q pixels. $I(x,y)$ is the grayscale value, $g(x,y)$ is the non-local mean filtered values, then i in the new histogram horizontal and vertical coordinates (i,j) denotes the grayscale values, and j denotes the non-local mean filtered value. Meanwhile, the size of the original image and the size of the generated non-local mean filtered image are kept the same; therefore, the corresponding non-local mean filtered 2D histogram can be generated from the non-local mean image and the grayscale image. Furthermore, by normalization process, the following equation shows:

$$P_{ij} = \frac{h(i,j)}{M \times N} \tag{13}$$

The final 2D histogram can be formed. **Figure 2** shows the three-dimensional (3D) views of 2D histograms formed by normalizing it with Equation (11).

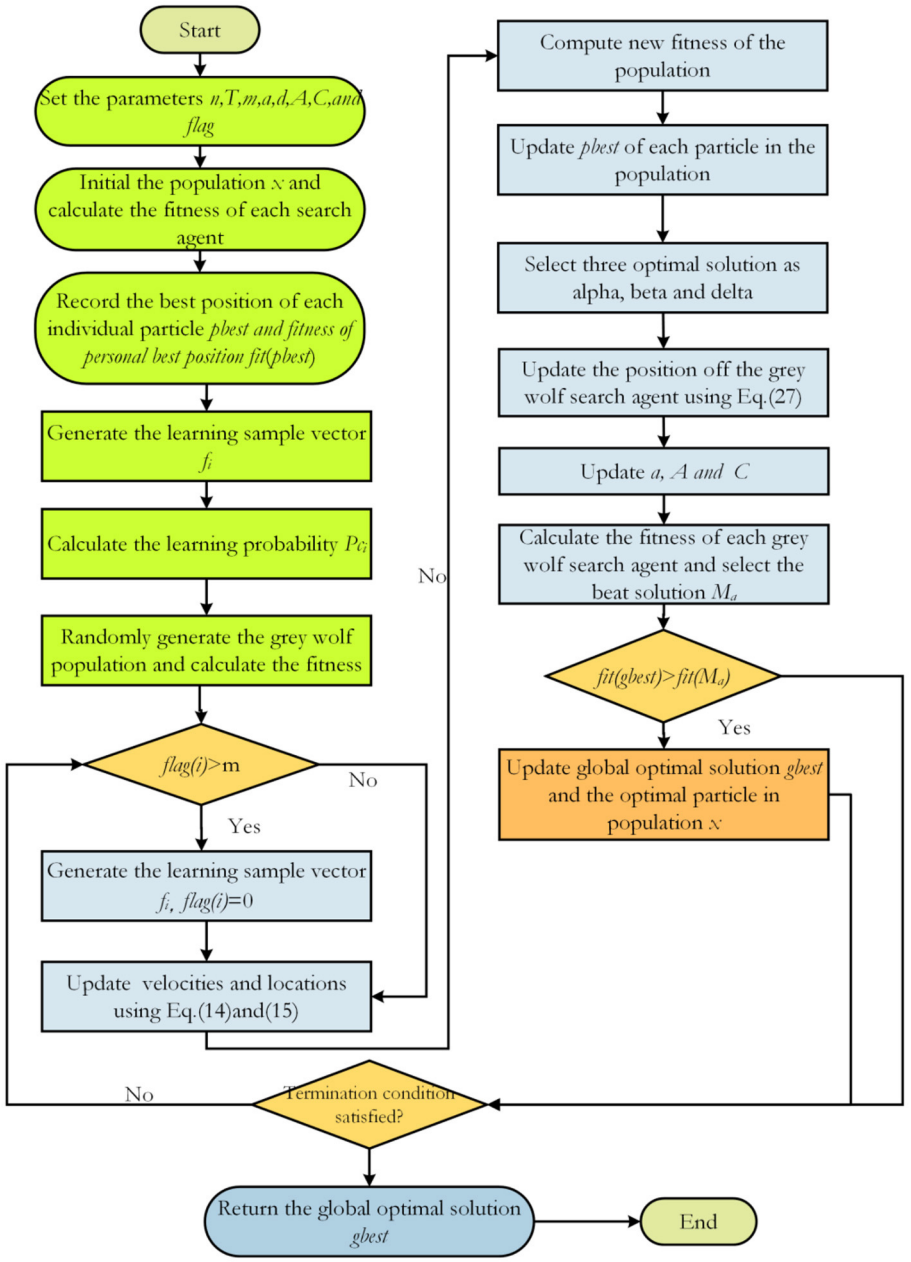


FIGURE 3 | Flowchart of GCLPSO.

ENHANCED COMPREHENSIVE LEARNING PARTICLE SWARM OPTIMIZER

There have been many studies on swarm intelligence optimization methods for finding optimal thresholds of images. In this study, we refer to a recently improved comprehensive particle swarm algorithm, GCLPSO, which improves the exploration and detection capability of CLPSO and improves the ability to find the optimal threshold and this study presents this algorithm.

Comprehensive Learning Particle Swarm Optimizer

The CLPSO algorithm was proposed by Liang et al. (2006). It uses a new comprehensive learning strategy (CLS) to update the velocity of particles using the personal best position *pbest* of other particles. CLS can maintain the diversity of the population and prevent premature maturation. The velocity and position update formula in the CLPSO algorithm is shown as follows:

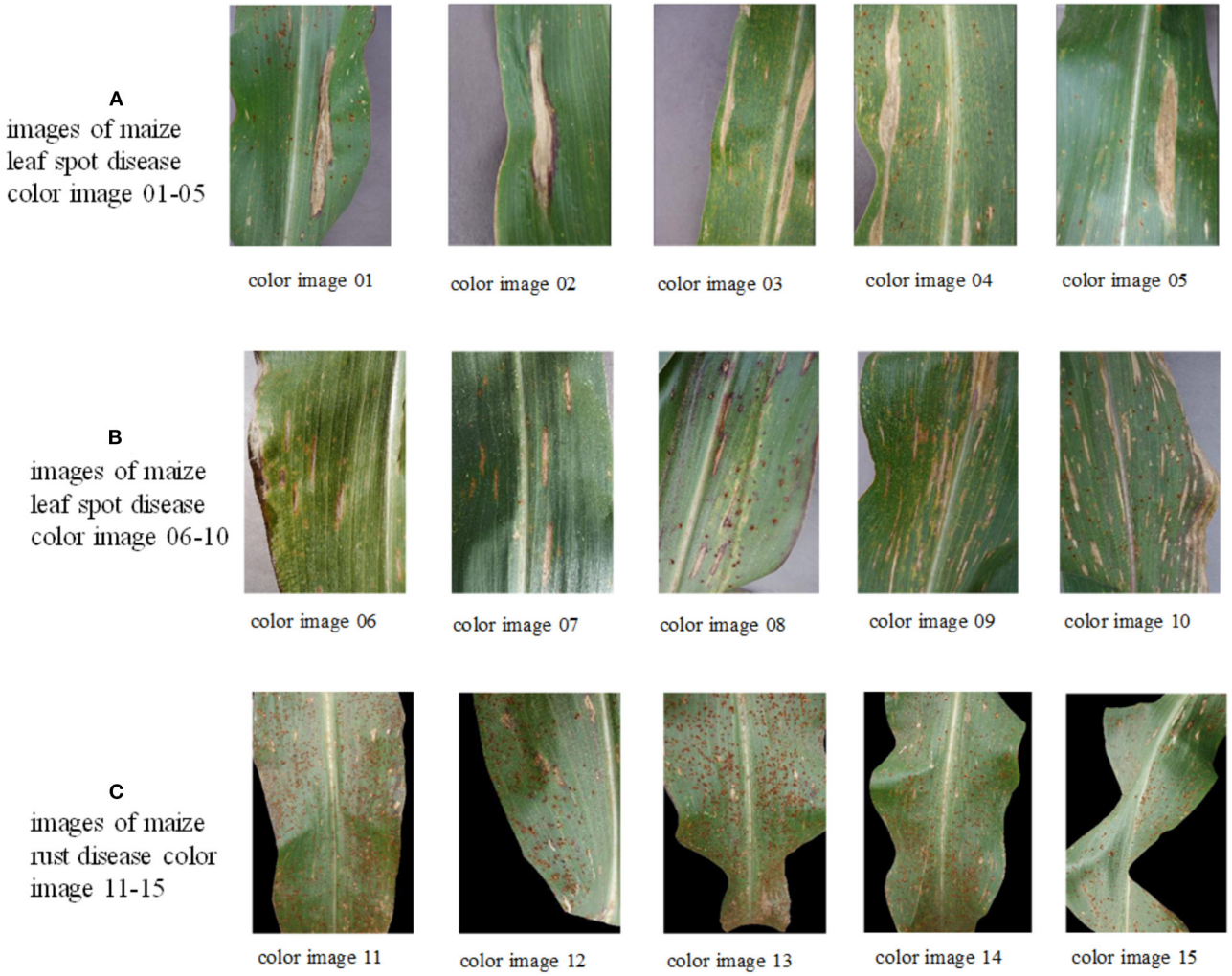


FIGURE 4 | Original images of maize leaf spot disease color images 1–5, maize gray spot disease color images 6–10, and maize rust disease color images 11–15 in the Plant Village.

$$v_{id} = w*v_{id} + c*r_{id}\left(pbest_{f_i(d),d} - x_{id}\right) \quad (14)$$

$$x_{id} = x_{id} + v_{id} \quad (15)$$

where $f_i(d)$ denotes the dimension value of the dth dimension in a particle $pbest$, $f_i = [f_i(1), f_i(2), \ldots, f_i(D)]$ denotes the learning sample vector defined for particle i,$pbest_{f_i(d),d}$ denotes the best position of the particle among all the particles $pbest$ corresponding dimensional value. The dimension of which particle is learned depends on the parameter learning probability Pc. For each dimension of a particle, we generate a random number. If this random number is greater than Pc, the corresponding dimension will be learned from its own $pbest$. Otherwise, it will be learned in the other particle $pbest$. The algorithm selects learning particles from other particles as follows:

(1) First, select two particles from the population at random, excluding the velocity update particles.
(2) Then, compare the fitness values of the $pbest$ of these two particles and choose the better one. In this study, the fitness value is the minimum solution of the function, which means that the smaller the function value is, the better when solving the minimization problem.

The CLPSO first assigns the learning probability Pc to each particle using the following equation:

$$Pc_i = a + b*\frac{\exp\left(\frac{10*(i-1)}{N-1}\right)}{\exp(10) - 1} \quad (16)$$

where a and b are two parameters used to identify the maximum and minimum learning probabilities and N is the total number of particles.

In addition to avoid wasting time in undesirable directions when the particle learns the best position of the particle individual from the sample, a particle learning count threshold m is defined and if the adaptation value of the particle does not improve after m consecutive moves, a random particle is generated again instead of the particle.

Gray Wolf Optimizer

Mirjalili et al. (2014) proposed the metaheuristic algorithm grey wolf optimization (GWO) in 2014, a variant of the PSO with a metaphor, as proven in the recent works (Villalón et al., 2020). Similar to other metaheuristic approaches (Ala et al., 2020; Seifi et al., 2020; Moayedi and Mosavi, 2021a,b), the algorithm is inspired by the social hierarchy and hunting strategies of gray wildlife wolves and it has been applied to various problems due to its simple idea (Heidari and Pahlavani, 2017; Aljarah et al., 2019; Heidari et al., 2019; Tang et al., 2020). Regardless of its defect, we still can see some performance features in this method (Niu et al., 2019; Hu et al., 2021). In this class of methods, an initial set of agents needs to be evolved to increase the capacity to explore trends of the method within the searching process (Moayedi and Mosavi, 2021c). The best agents should be considered as alpha (α), beta (β), and delta (δ) to help other agents omega (ω) to explore more favorable areas of solution space.

In GWO, agents can identify the location of their prey and surround them. To mathematically model this behavior, the equation is as follows:

$$\overrightarrow{D} = \left|\overrightarrow{C} * \overrightarrow{X}_p(t) - \overrightarrow{X}(t)\right| \tag{17}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X}_p(t) - \overrightarrow{A} * \overrightarrow{D} \tag{18}$$

where t is the number of iterations, $\overrightarrow{A}$ and $\overrightarrow{C}$ are the coefficient vectors, $\overrightarrow{X}_p$ is the position vector of the prey, and $\overrightarrow{X}$ is the position vector of the gray wolf.

$\overrightarrow{A}$ and $\overrightarrow{C}$ are calculated as shown below:

$$\overrightarrow{A} = 2\overrightarrow{a} * \overrightarrow{r}_1 - \overrightarrow{a} \tag{19}$$

$$\overrightarrow{C} = 2\overrightarrow{r}_2 \tag{20}$$

where $\overrightarrow{a}$ is decreasing from 2 to 0 with increasing number of iterations, $\overrightarrow{r}_1$和$\overrightarrow{r}_2$ are random numbers between 0 and 1.

Alpha (α) agents usually lead the hunting process. So, the behavior is described by the following equation (Chantar et al., 2020):

$$\overrightarrow{D}_\alpha = \left|\overrightarrow{C}_1 * \overrightarrow{X}_\alpha - \overrightarrow{X}\right| \tag{21}$$

$$\overrightarrow{D}_\beta = \left|\overrightarrow{C}_2 * \overrightarrow{X}_\beta - \overrightarrow{X}\right| \tag{22}$$

$$\overrightarrow{D}_\delta = \left|\overrightarrow{C}_3 * \overrightarrow{X}_\delta - \overrightarrow{X}\right| \tag{23}$$

$$\overrightarrow{X}_1 = \overrightarrow{X}_\alpha - \overrightarrow{A}_1 * \left(\overrightarrow{D}_\alpha\right) \tag{24}$$

$$\overrightarrow{X}_2 = \overrightarrow{X}_\beta - \overrightarrow{A}_2 * \left(\overrightarrow{D}_\beta\right) \tag{25}$$

$$\overrightarrow{X}_3 = \overrightarrow{X}_\delta - \overrightarrow{A}_3 * \left(\overrightarrow{D}_\delta\right) \tag{26}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X}_1 + \overrightarrow{X}_2 + \overrightarrow{X}_3}{3} \tag{27}$$

Enhanced CLPSO

The CLPSO is a well-known variant of the PSO (Fan et al., 2022), which updates the velocity of the particles by the *pbest* of all the particles, which prevents the algorithm from falling into a local optimum prematurely and prevents the algorithm from performing a local search near the global optimum. The improved algorithm GCLPSO first selects the optimal three solutions of the CLPSO algorithm, as the gray wolf algorithm alpha (α), beta (β), and delta (δ). The optimal solution of each iteration in the CLPSO algorithm is searched nearby by the GWO idea, while the searched optimal solution is substituted for the optimal solution in the CLPSO algorithm. The specific procedures of the algorithm are described as follows:

(1) First, initialize the particles and parameters and calculate the fitness value of each particle
(2) Update each particle using the CLPSO algorithm
(3) The three optimal solutions in the CLPSO algorithm are selected as the gray wolf algorithm alpha (α), beta (β), delta (δ), and their optimal solutions are searched locally using the GWO algorithm nearby. If the optimal solution search is better than the optimal solution in the CLPSO algorithm, the optimal solution in the CLPSO algorithm will be replaced.
(4) Keep looping (2), (3) steps until the termination condition is satisfied.

The flowchart of the GCLPSO algorithm is shown in **Figure 3**.

In the GCLPSO algorithm, n is the size of the population, d is the dimensionality, g is the maximum number of iterations, the population initialization is O (n), the gray wolf population initialization is O (n), the update search particle position is O ($n \times d \times g$), the update local search of all the gray wolf positions is O ($n \times d \times g$), and the sorted population fitness value is O ($n \times \log n \times g$). Therefore, the final time complexity of the GCLPSO algorithm is 2O ($n \times d \times g + n$) + O ($n \times \log n \times g$).

EXPERIMENTS AND RESULTS

Image and Parameter Settings

The experiments in this study were conducted through the maize disease image library in the Plant Village public dataset, containing images of three kinds of maize diseases: maize leaf spot, maize gray spot, and maize rust diseases and five images of each disease were taken separately, with an image size of 256 × 512 pixels. They are shown separately in **Figure 4** below in (a) maize leaf spot, (b) maize gray spot, and (c) maize rust diseases.

All the experiments were conducted on a computer with a 3.40 GHz Intel® Core i7 processor and 16 GB of Random

TABLE 2 | Results of the feature similarity FSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize leaf spot disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	1/0/4	1/0/4	3/0/2	5/0/0	5/0/0	5/0/0
	Mean	**1.2**	2	3	5.2	8.2	9.8	7.8
	Rank	**1**	2	3	5	6	7	5
3	+/-/=	∼	1/0/4	2/1/2	5/0/0	5/0/0	5/0/0	5/0/0
	Mean	**1.6**	2.6	3.4	5.8	8.8	9.6	8.4
	Rank	**1**	2	3	4	6	7	5
4	+/-/=	∼	0/0/5	1/0/4	3/0/2	4/0/1	5/0/0	2/0/3
	Mean	**2.2**	5.2	4.6	5.8	9.2	9.8	7
	Rank	**1**	3	2	4	6	7	5

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

TABLE 3 | Results of the peak signal-to-noise ratio (PSNR) comparison by the Wilcoxon signed-rank test at each threshold level of maize leaf spot disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	0/0/5	1/0/4	3/0/2	4/0/1	5/0/0	3/0/2
	Mean	**1.4**	2	2.6	5.6	8.6	8.4	7.2
	Rank	**1**	2	3	4	7	6	5
3	+/-/=	∼	1/0/4	1/0/4	4/0/1	5/0/0	5/0/0	5/0/0
	Mean	**1.8**	2.8	3.4	5.2	8.8	8.6	8.4
	Rank	**1**	2	3	4	7	6	5
4	+/-/=	∼	1/0/4	1/0/4	3/0/2	4/0/1	5/0/0	3/0/2
	Mean	**2.4**	4.4	5.6	6	9	9.4	7
	Rank	**1**	2	3	4	6	7	5

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

TABLE 4 | Results of the structural similarity (SSIM) comparison by the Wilcoxon signed-rank test at each threshold level of maize leaf spot disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	0/0/5	2/0/3	3/0/2	4/0/1	5/0/0	3/0/2
	Mean	**1.2**	2	3	5.2	8.8	7.8	7
	Rank	**1**	2	3	4	8	7	5
3	+/-/=	∼	0/0/5	1/0/4	4/0/1	5/0/0	5/0/0	5/0/0
	Mean	**1.8**	3.2	3.2	4.8	8.2	8.6	7.8
	Rank	**1**	2	2	3	5	6	4
4	+/-/=	∼	1/0/4	1/0/4	2/0/3	4/0/1	5/0/0	2/0/3
	Mean	**2.6**	5.6	4.4	5.2	9	9.4	6.8
	Rank	**1**	4	2	3	6	7	5

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

Access Memory (RAM) and programming was performed using MATLAB 2018b.

In this section of experiments, GCLPSO will be used for the practical application of multi-threshold maize disease image segmentation. To verify the effectiveness of GCLPSO on multi-threshold image segmentation, GCLPSO will be compared with CLPSO, two improved algorithms SCADE and m_SCA, and three original algorithms SSA, SCA, and SMA, respectively. To ensure the validity and fairness of the experiments (Chen et al., 2021; Moayedi and Mosavi, 2021d; Nosratabadi et al., 2021; Yang et al., 2021), all the algorithms involved in the comparisons were conducted under the same experimental conditions. Such a setting is one of the most crucial rules in the artificial intelligence community (Song et al., 2020; Thaher et al., 2020; Mousavi et al., 2021; Tavoosi et al., 2021). The population size was set to 20, the maximum number of evaluations *MaxFEs* was uniformly set to 100, and all the algorithms were tested

TABLE 5 | Results of the FSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize gray spot disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	0/0/5	1/0/4	4/0/1	5/0/0	5/0/0	5/0/0
	Mean	**1.4**	2	3.2	5.6	6.4	10	9
	Rank	**1**	2	3	4	5	7	6
3	+/-/=	∼	1/0/4	1/0/4	2/0/3	5/0/0	5/0/0	5/0/0
	Mean	**2.2**	3.4	2.2	6	9.4	9.2	7.6
	Rank	**1**	2	1	3	6	5	4
4	+/-/=	∼	1/0/4	1/0/4	2/0/3	5/0/0	5/0/0	3/0/2
	Mean	**3**	5.4	4	6.8	9.2	9.8	6.8
	Rank	**1**	3	2	4	5	6	4

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

TABLE 6 | Results of the PSNR comparison by the Wilcoxon signed-rank test at each threshold level of maize gray spot disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	0/0/5	1/0/4	3/0/2	3/0/2	5/0/0	5/0/0
	Mean	**1.4**	2.4	3.4	5.2	6.8	9.4	9.2
	Rank	**1**	2	3	4	5	7	6
3	+/-/=	∼	1/0/4	2/0/3	2/0/3	4/0/1	3/0/2	3/0/2
	Mean	**2.6**	3	3.6	5.6	9.6	8.6	7.2
	Rank	**1**	2	3	4	8	7	5
4	+/-/=	∼	1/0/4	1/0/4	1/0/4	3/0/2	2/0/3	2/0/3
	Mean	**3.4**	3.8	4.8	5.8	9.4	9	6.6
	Rank	**1**	2	3	4	7	6	5

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

TABLE 7 | Results of the SSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize gray spot disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	0/0/5	1/0/4	4/0/1	4/0/1	5/0/0	5/0/0
	Mean	**1.8**	2.6	2.8	5.2	7.2	9.4	9.2
	Rank	**1**	2	3	4	5	7	6
3	+/-/=	∼	1/0/4	2/0/3	2/0/3	4/0/1	5/0/0	4/0/1
	Mean	**2.4**	3.2	3	5.4	9.6	8.8	7.4
	Rank	**1**	3	2	4	7	6	5
4	+/-/=	∼	0/0/5	1/0/4	1/0/4	4/0/1	3/0/2	2/0/3
	Mean	**3.6**	4	4.4	6.4	9.4	9.4	6.4
	Rank	**1**	2	3	4	5	5	4

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

30 times independently to reduce the influence of random conditions. The thresholds of the experiments were set to 2, 3, and 4 thresholds, respectively. Meanwhile, the segmentation results were evaluated FSIM, PSNR, and SSIM to compare the image segmentation results, respectively. Also, we further analyzed the evaluation results of FSIM, PSNR, and SSIM using mean, variance, and the Wilcoxon signed-rank test (Liu et al., 2021a).

Evaluation Indicators

The PSNR, SSIM index, and FSIM index is applied to further evaluate image segmentation quality (Liu et al., 2021b; Shi et al., 2021; Zhang et al., 2021).

The FSIM represents the FSIM between the original image and the segmented image. FSIM is composed of high phase composite (PC) and gradient amplitude (GM) to evaluate local structure and provide contrast information. Its value range is between 0 and

TABLE 8 | Results of the FSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize rust disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	0/1/4	1/0/4	1/0/4	1/0/4	4/0/1	4/0/1
	Mean	**3.6**	3.8	4.2	5.6	5.8	9.2	9
	Rank	**1**	2	3	4	5	7	6
3	+/-/=	∼	2/0/3	4/0/1	3/0/2	5/0/0	5/0/0	5/0/0
	Mean	**1**	3.2	6.4	4.4	8.6	9.2	8.6
	Rank	**1**	2	4	3	5	6	5
4	+/-/=	∼	1/0/4	2/0/3	2/0/3	4/0/1	5/0/0	4/0/1
	Mean	**2**	5.6	6.8	3.8	8.4	9.4	7.8
	Rank	**1**	3	4	2	6	7	5

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

TABLE 9 | Results of the PSNR comparison by the Wilcoxon signed-rank test at each threshold level of maize rust disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	**∼**	0/1/4	1/1/3	1/0/4	1/0/4	4/0/1	4/0/1
	Mean	**4.4**	4.4	3.6	5.6	5.6	9.6	8.8
	Rank	**1**	1	1	3	3	5	4
3	+/-/=	∼	1/0/4	4/0/1	3/0/2	5/0/0	5/0/0	5/0/0
	Mean	**1.2**	3.2	6.6	4.8	8.6	9.2	8.6
	Rank	**1**	2	4	3	5	6	5
4	+/-/=	∼	1/0/4	2/0/3	2/0/3	4/0/1	5/0/0	5/0/0
	Mean	**1.6**	6	6.2	4.2	7.8	9.4	8
	Rank	**1**	3	4	2	5	7	6

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

TABLE 10 | Results of the SSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize rust disease.

Thresholds		GCLPSO	CLPSO	SCADE	m_SCA	SSA	SCA	SMA
2	+/-/=	∼	0/1/4	1/0/4	1/0/4	1/0/4	5/0/0	4/0/1
	Mean	**4.4**	4.4	4.8	5.2	5.6	9.6	8.8
	Rank	**1**	1	2	3	4	6	5
3	+/-/=	∼	1/0/4	4/0/1	3/0/2	5/0/0	5/0/0	5/0/0
	Mean	**1.2**	3	6.4	5	8.4	9.2	8.6
	Rank	**1**	2	4	3	5	7	6
4	+/-/=	∼	2/0/3	2/0/3	2/0/3	4/0/1	5/0/0	4/0/1
	Mean	**1.6**	6.2	6.2	4	8	9.4	8
	Rank	**1**	3	3	2	4	5	4

Mean, means the mean value of the ranking of the SSIM evaluation results obtained by the algorithm after segmentation of each image at the 2 threshold level, Rank, means the overall ranking.

1, where the closer to 1, the better the segmentation effect. The detailed expression is as follows:

$$FSIM = \frac{\sum_{I\in\Omega} S_L(X)\, PC_m(X)}{\sum_{I\in\Omega} PC_m(X)} \tag{28}$$

$$S_L(X) = S_{PC}(X)\, S_G(X) \tag{29}$$

$$S_{PC}(X) = \frac{2PC_1(X)\, PC_2(X) + T_1}{PC_1^2(X)\, PC_2^2(X) + T_1} \tag{30}$$

$$S_G(X) = \frac{2G_1(X)\, G_2(X) + T_2}{G_1^2(X)\, G_2^2(X) + T_2} \tag{31}$$

$$G = \sqrt{G_x^2 + G_y^2} \tag{32}$$

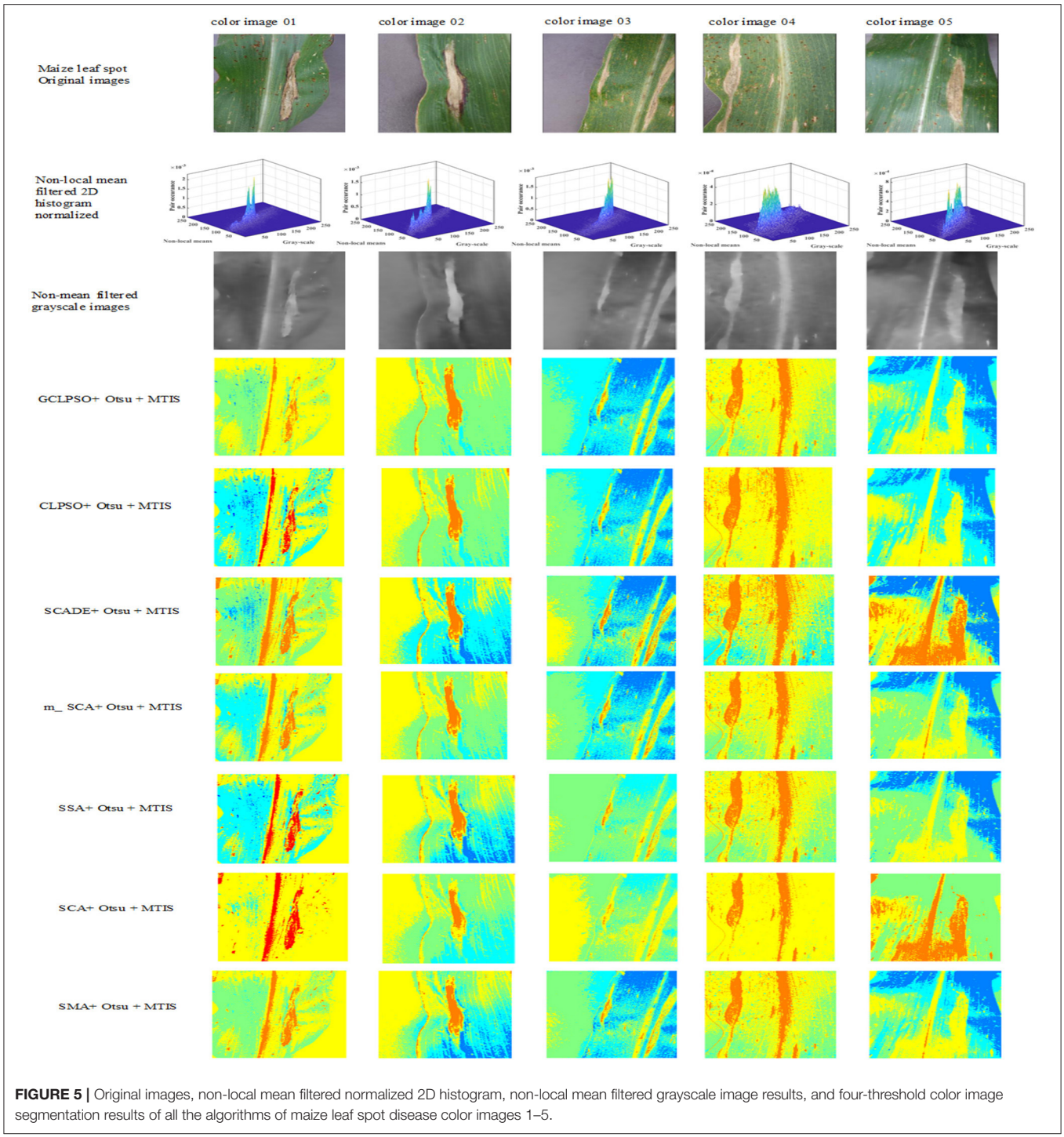


FIGURE 5 | Original images, non-local mean filtered normalized 2D histogram, non-local mean filtered grayscale image results, and four-threshold color image segmentation results of all the algorithms of maize leaf spot disease color images 1–5.

$$PC(X) = \frac{E(X)}{\left(\varepsilon + \sum_m A_n(X)\right)} \tag{33}$$

The Ω denotes all the pixel domains of the original image. $S(X)$ denotes the similarity score. $PC(x)$ means the phase consistency measure, T_1 and T_2 are constants, G is the gradient descent method, $E(X)$ response vector size position X and scale n, furthermore, ε is a small quantity and $A_n(X)$ is a local scale size n.

The PSNR indicates the difference between the segmented image and the original image. The larger the value, the better.

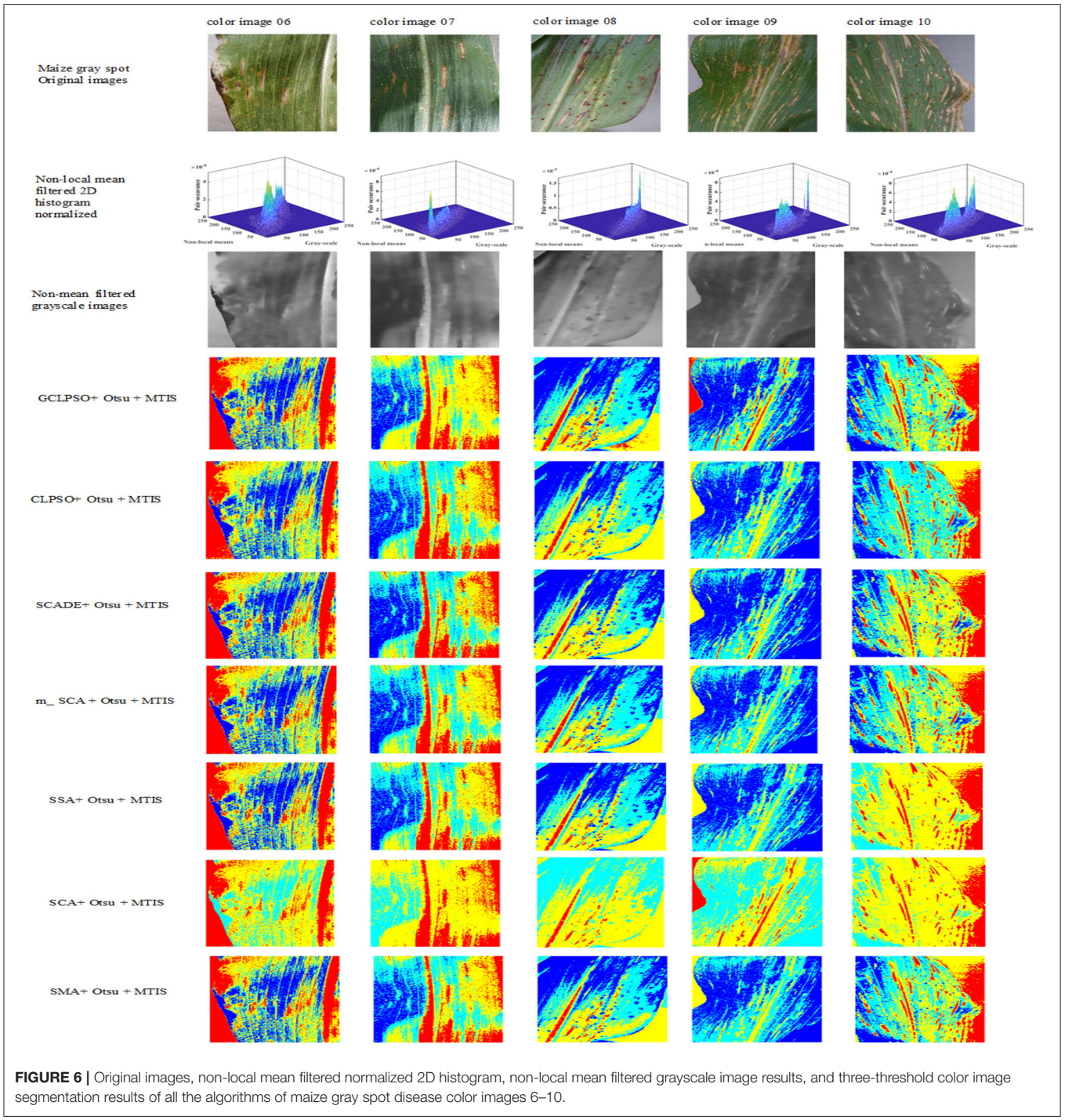


FIGURE 6 | Original images, non-local mean filtered normalized 2D histogram, non-local mean filtered grayscale image results, and three-threshold color image segmentation results of all the algorithms of maize gray spot disease color images 6–10.

It is defined as shown in Equation (34).

$$PSNR = 20 \cdot \left(\frac{255}{RMSE}\right) \tag{34}$$

The SSIM represents a measure of the similarity of two images. The greater its value, the more effective is the segmentation of the threshold. Its definition is illustrated by Equation (35).

$$SSIM = \frac{\left(2\mu_I\mu_{Seg} + c_1\right)\left(2\sigma_{I,Seg} + c_2\right)}{\left(\mu_I{}^2 + \mu_{Seg}{}^2 + c_1\right)\left(\sigma_I{}^2 + \sigma_{Seg}{}^2 + c_2\right)} \tag{35}$$

μ_I, μ_s, σ_I, and σ_s are the mean values and SD of the original and segmented images, respectively, $_{Seg}$ is the covariance of

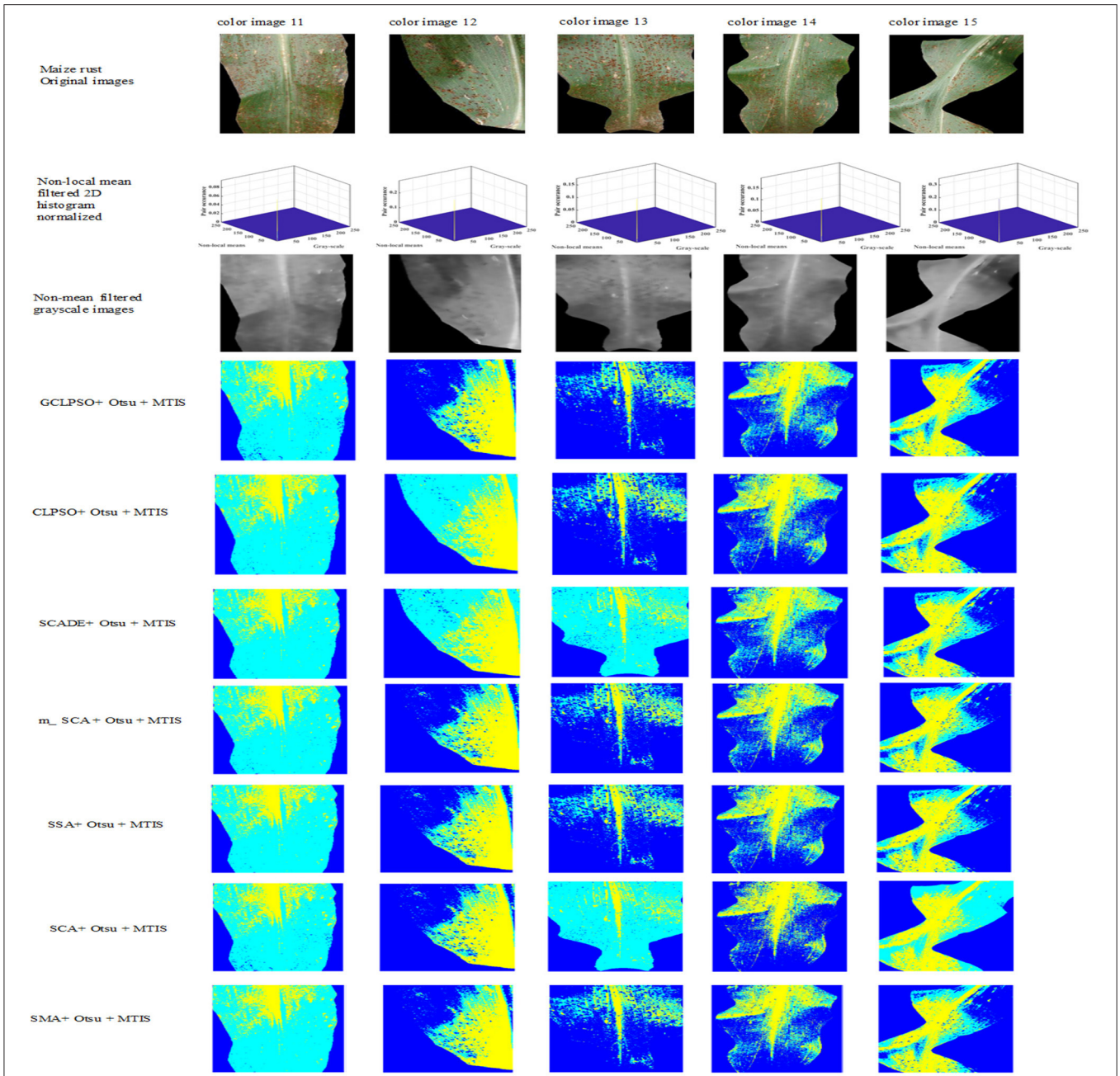


FIGURE 7 | Original images, non-local mean filtered 2D histogram, non-local mean filtered grayscale image results, and two-threshold color image segmentation results of all the algorithms of maize rust disease color images 11–15.

the basic image and the image segmentation, and c_1 and c_2 represent constants.

Experimental Results and Analysis

In this study, in order to evaluate the image segmentation effect of GCLPSO at a multi-threshold, the comparison algorithms involved are CLPSO, SCADE, m_ SCA, SSA, SCA, and SMA. The segmentation results are evaluated using the PSNR, SSIM, and FSIM and the results are evaluated by the ways of mean, variance, and the Wilcoxon signed-rank test. **Tables 2–4** show the results of the FSIM, PSNR, and SSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize leaf spot disease; **Tables 5–7** exhibit the results of the FSIM, PSNR, and SSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize gray spot disease; **Tables 8–10** show the results of the FSIM, PSNR, and SSIM comparison by the Wilcoxon signed-rank test at each threshold level of maize rust disease. Furthermore, mean means average ranking, rank means ranking

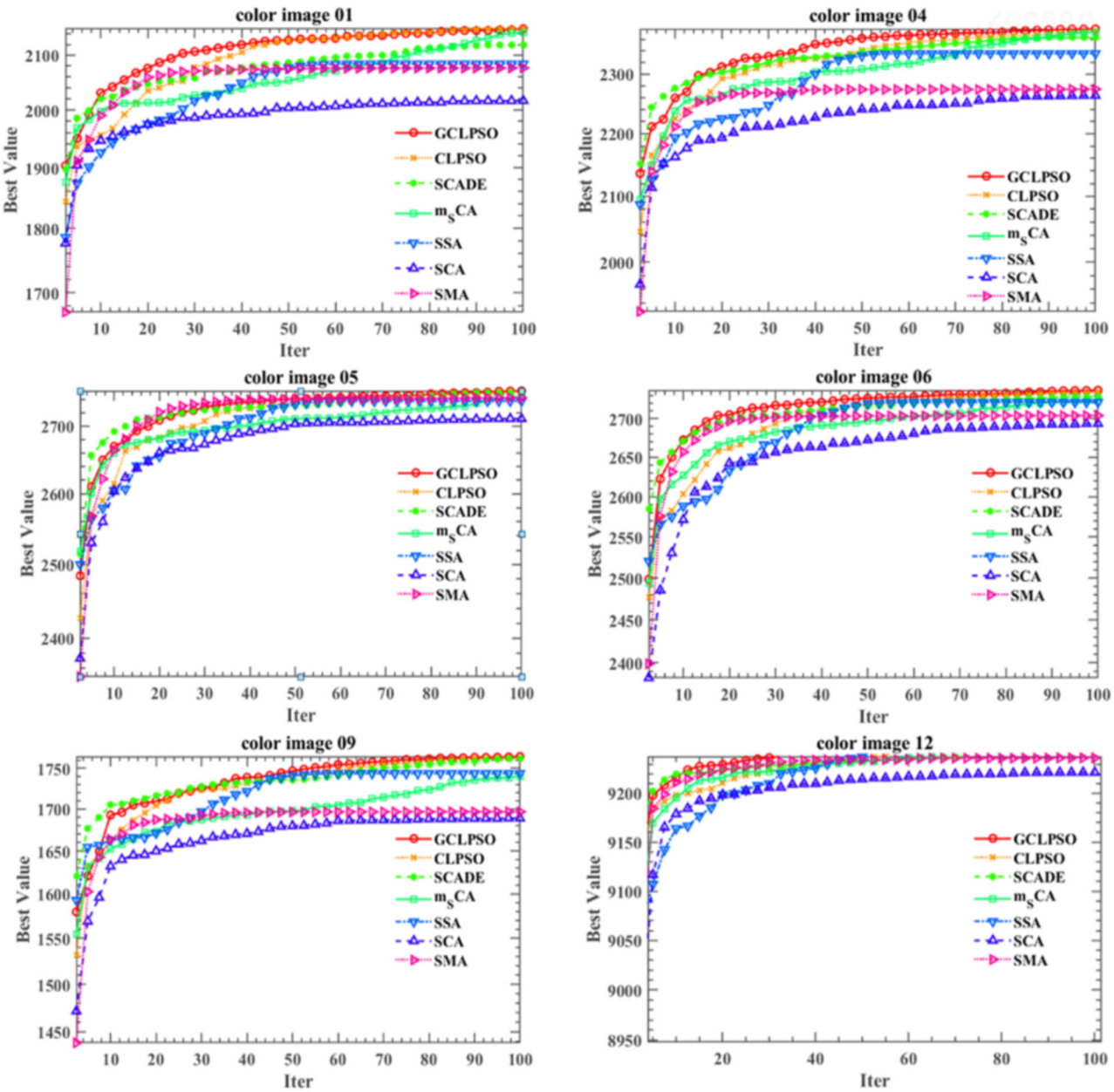


FIGURE 8 | Convergence curves of 2D-Otsu at threshold 2.

order, "+" means that the performance of GCLPSO is better than the comparison algorithm, "=" means that the performance of GCLPSO is equal to the comparison algorithm, and "-" means that the performance of GCLPSO is worse than the comparison algorithm. It can be seen that GCLPSO ranks first overall and has the best performance.

In the experience on MTIS, images 1–5 are maize leaf spot disease, images 6–10 are maize gray spot disease, and images 11–15 are maize rust disease. The image size was set to 256 × 512 pixels. **Figure 5** shows that the first row presents the original images of five randomly selected color images of corn leaf spot disease in the Plant village public dataset; the second row presents the grayscale and non-mean filtered normalized 3D histogram of the corresponding image, with the X-axis as the gray value of the grayscale image, the Y-axis as the gray value of the non-mean filtered image and the Z-axis as the normalized result of the combination of the gray value of the grayscale image and the non-mean filtered image; the third row is the non-local mean filtered grayscale image of the corresponding image; the remaining rows are the color segmentation results

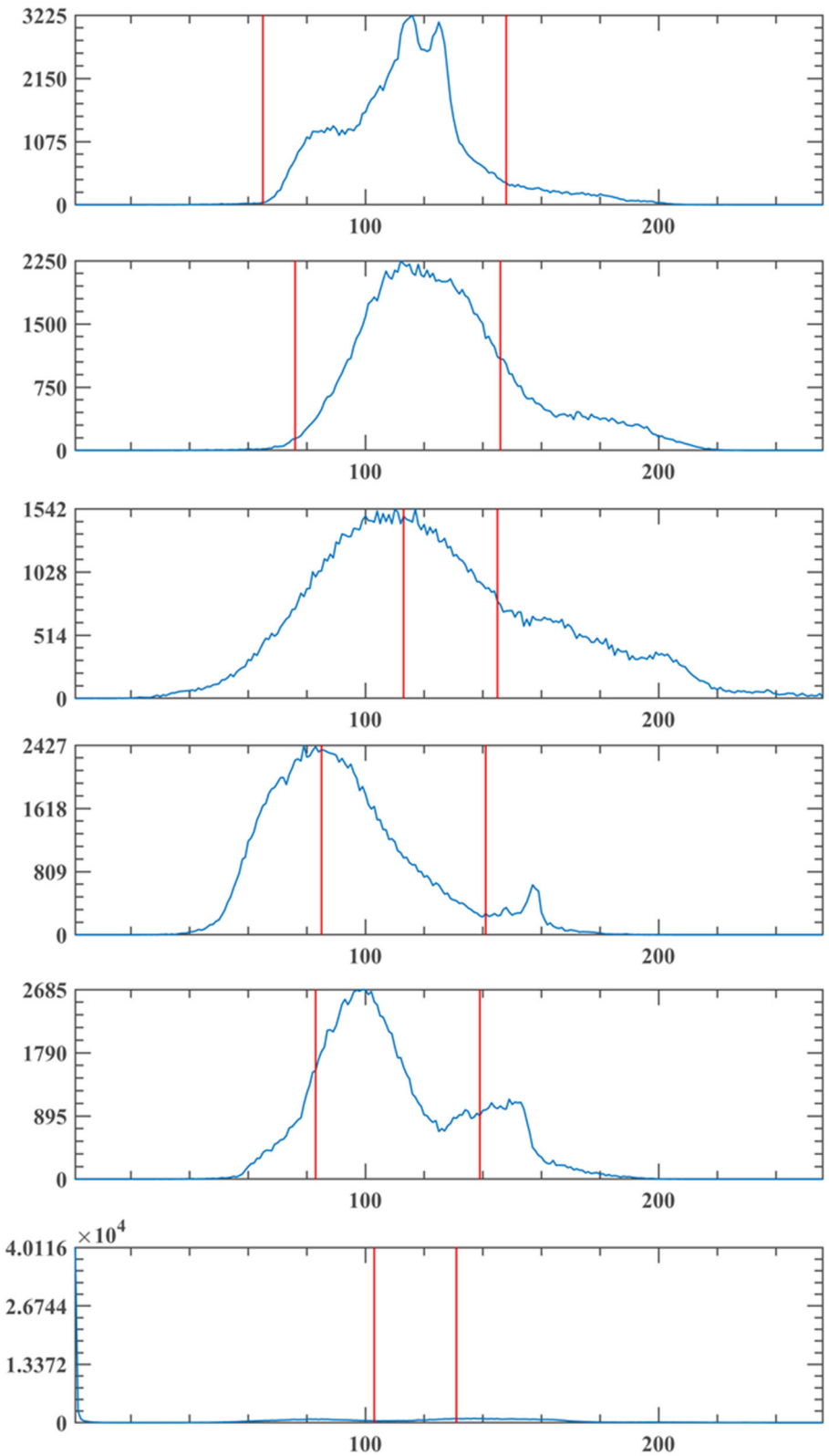


FIGURE 9 | Segmentation threshold line of GCLPSO-2D-Otsu at threshold 2.

of the GCLPSO algorithm combined with the non-local mean filter and the contrast algorithm to optimize the Otsu segmentation with four-threshold, respectively. **Figure 6** denotes the original images; non-local mean filtered normalized 2D histogram, non-local mean filtered grayscale image results, and three-threshold color image segmentation results of all the maize gray spot disease algorithms—color images 6–10. **Figure 7** is similar to **Figures 5**, **6** for the two-threshold color segmentation results of all the corn rust spot disease algorithms—color images 11–15. We can see that the segmentation effect under the GCLPSO algorithm is significantly better than other similar algorithms.

According to **Figure 8**, the local features of the image acquired by GCLPSO are obviously better than those of other algorithms. It is easy to conclude that GCLPSO surpasses other competing algorithms at the threshold level of 2.

Conclusion

Figures 5–**7** show the segmentation results under different algorithms; from the segmentation results, the local features of the images retained by GCLPSO are better than those retained by other algorithms. From the data, **Tables 2**–**10** show the comparison of mean and SD of FSIM index, PSNR, and SSIM index for all the algorithms segmentation of maize leaf spot, gray spot, and rust images. **Appendix Tables 1–9** show the results of FSIM index, PSNR, and SSIM index comparison by the Wilcoxon signed-rank test at each threshold level for maize leaf spot, gray spot, and rust images. The comparison results showed that the feature similarity index outperformed other algorithms at thresholds 2, 3, and 4 for all the disease images and also performed significantly better than other algorithms at other thresholds. **Figure 8** shows the convergence curves of images 1, 4, 5, 6, 9, and 12 when the image segmentation experiments were performed at a threshold level of 2, 3, and 4. Based on the convergence curves, it can be seen that GCLPSO is better at finding the maximum value of the maximum interclass variance and it has higher convergence accuracy than other algorithms. Therefore, based on the above comparison and experimental analysis, GCLPSO outperformed the other algorithms at the level of multiple thresholds for maize with multiple disease spots. The threshold values and segmentation threshold lines selected by GCLPSO for the 2-threshold segmentation of maize disease image (01)(04)(06)(09)(10)(12) can be seen in **Figure 9** above, and it can be seen that the threshold values were selected more reasonably.

SUMMARY AND FUTURE WORK

According to the multi-threshold image segmentation experiments on three maize disease images, the used GCLPSO was compared with CLPSO and two other improved algorithms and three original algorithms, respectively, and the multi-threshold segmentation comparison experiments were compared with each other at the same time. In addition, the PSNR, SSIM, and FSIM were used to compare the image segmentation results for evaluation and the evaluation results were analyzed using the mean, variance, and the Wilcoxon signed-rank tests. It can be seen from the experiments in this study that the GCLPSO algorithm can be used as an optimization tool for the Otsu segmentation of maize disease images in multi-threshold level segmentation results by non-local average filtered 2D histogram using an improved swarm intelligent optimization algorithm since GCLPSO can obtain the best selection of thresholds and has ideal stability in the segmentation process. Therefore, it can be effectively used in maize leaf disease image segmentation.

In future research work, the GCLPSO can be combined with other optimization methods for image segmentation of multicrop diseases to improve the identification and intelligent diagnosis of disease deficiency in maize and other crops and effectively reduce the economic losses caused by crop diseases.

(a) Maize leaf spot disease

Maize leaf spot disease, also known as stripe disease, coal stripe disease, leaf blight disease, and large spot disease, is a major foliar disease of maize, which occurs throughout China and causes heavy damage. It mainly affects the leaves, the leaf sheaths, and bracts in severe cases. Generally, from the bottom of the leaf first, it gradually expands upward, when serious spots can spread throughout the plant, but there are also from the upper leaves of the case. Infected leaves form large nucleate spots, which are initially water-stained greenish-gray or grayish-green spots in the field and then expand into large diamond-shaped or long fusiform spots with dark brown margins and light brown or gray centers, generally 5–10 cm long, with a distinct black-brown mold layer on the spots when wet, and in severe cases, the spots combine to split longitudinally, and the leaves die.

(b) Maize gray spot disease

Maize gray spot disease, also known as tail spore leaf spot disease and corn mold spot disease, is one of the diseases that have been rising rapidly and causing more serious damage in the recent years. It mainly affects the leaves. At first, on the leaf surface to form no obvious edge of the oval, moment round gray to light brown spots, later turned brown. The spots are mostly limited to between the parallel leaf veins, size 4–20 × 2–5 (mm). When the humidity is high, the back of the spot produces gray moldy material, i.e., the conidiophore and conidia of the disease. Sometimes, it causes the fruiting spike to rot or droop and the seeds to blacken.

(c) Maize rust spot disease

Maize rust mainly affects the leaves, but it can also occur on the cob bracts and male flowers in severe cases. The upper and middle leaves of the plant are heavily affected, initially with inconspicuous yellowish dots scattered or clustered on the adaxial surface of the leaves, later protruding and expanding into a round to oblong, yellowish brown or brown, with the surrounding epidermis turning up and scattering rust-colored powder (summer spores of the pathogenic fungus). Later on, the spot grows round black protrusions, rupture to reveal black-brown powder (winter spores of the pathogenic fungus). The disease is caused by the fungus, the field leaf disease, the disease produced by the summer spores spread by airflow, reinfestation, spread, and expansion. Production of early maturing varieties is susceptible to the disease, heavy incidence of biased nitrogen fertilization, high temperature, humidity, rain, foggy days, and

insufficient light facilitate the prevalence of maize rust (Lv et al., 2020).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. This data can be found here: https://github.com/Tmcsn/AI-Challenger-2018-CropDisease.

AUTHOR CONTRIBUTIONS

CC and AH: writing—original draft, writing—review and editing, software, visualization, and investigation. XW, HC, and HY: conceptualization, methodology, formal analysis, investigation, writing—review and editing, funding acquisition, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Science and Technology Development Project of Jilin Province (20190301024NY).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.789911/full#supplementary-material

REFERENCES

Abbassi, R., Abbassi, A., Heidari, A. A., and Mirjalili, S. (2019). An efficient salp swarm-inspired algorithm for parameters identification of photovoltaic cell models. *Energy Conver. Manage.* 179, 362–372. doi: 10.1016/j.enconman.2018.10.069

Ala, M., A.-Z., Heidari, A. A., Habib, M., Faris, H., Aljarah, I., et al. (2020). "Salp chain-based optimization of support vector machines and feature weighting for medical diagnostic information systems," in *Evolutionary Machine Learning Techniques* (Amman, Jordan, Springer). doi: 10.1007/978-981-32-9990-0_2

Ale, L., Sheta, A., Li, L., Wang, Y., and Zhang, N. (2019). "Deep learning based plant disease detection for smart agriculture," in *2019 Ieee Globecom Workshops*. (Waikoloa, HI, USA, IEEE). doi: 10.1109/GCWkshps45667.2019.9024439

Aljarah, I., Mafarja, M., Heidari, A. A., Faris, H., and Mirjalili, S. (2019). Clustering analysis using a novel locality-informed grey wolf-inspired clustering approach. *Knowl. Inf. Syst.* 62, 507–539. doi: 10.1007/s10115-019-01358-x

Aljarah, I., Mafarja, M., Heidari, A. A., Faris, H., Zhang, Y., and Mirjalili, S. (2018). Asynchronous accelerating multi-leader salp chains for feature selection. *Appl. Soft Comput.* 71, 964–979. doi: 10.1016/j.asoc.2018.07.040

Bao, X. L., Jia, H. M., and Lang, C. B. (2019). A novel hybrid harris hawks optimization for color image multilevel thresholding segmentation. *Ieee Access.* 7, 76529–76546. doi: 10.1109/ACCESS.2019.2921545

Chantar, H., Mafarja, M., Alsawalqah, H., Heidari, A. A., Aljarah, I., and Faris, H. (2020). Feature selection using binary grey wolf optimizer with elite-based crossover for Arabic text classification. *Neural Comput. App.* 32, 12201–12220. doi: 10.1007/s00521-019-04368-6

Chen, C. C., Wang, X. C., Yu, H. L., Zhao, N. N., Wang, M. J., and Chen, H. L. (2020). An enhanced comprehensive learning particle swarm optimizer with the elite-based dominance scheme. *Complexity*. 2020:4968063. doi: 10.1155/2020/4968063

Chen, Y., Chen, W., Chandra Pal, S., Saha, A., Chowdhuri, I., Adeli, B., et al. (2021). Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential. *Geocarto Int.* 2021, 1–21. doi: 10.1080/10106049.2021.1920635

Dhami, N. B., Kim, S. K., Paudel, A., Shrestha, J., and Rijal, T. R. (2015). A review on threat of gray leaf spot disease of maize in Asia. *J. Maize Res. Dev.* 1, 71–85. doi: 10.3126/jmrd.v1i1.14245

Elaziz, M. A., Heidari, A. A., Fujita, H., and Moayedi, H. (2020). A competitive chain-based Harris Hawks Optimizer for global optimization and multi-level image thresholding problems. *Appl. Soft Comput.* 2020:106347. doi: 10.1016/j.asoc.2020.106347

Fan, Y., Wang, P., Heidari, A. A., Chen, H., and Mafarja, M. (2022). Random reselection particle swarm optimization for optimal design of solar photovoltaic modules. *Energy* 239:121865. doi: 10.1016/j.energy.2021.121865

Faris, H., Mafarja, M. M., Heidari, A. A., Aljarah, I., Al-Zoubi, A. M., Mirjalili, S., et al. (2018). An efficient binary Salp Swarm Algorithm with crossover scheme for feature selection problems. *Knowledge-Based Syst.* 154, 43–67. doi: 10.1016/j.knosys.2018.05.009

Gao, L., and Lin, X. (2019). Fully automatic segmentation method for medicinal plant leaf images in complex background. *Comp. Electr. Agri.* 164:104924. doi: 10.1016/j.compag.2019.104924

Garcia, S., Fernandez, A., Luengo, J., and Herrera, F. (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Inf. Sci.* 180, 2044–2064. doi: 10.1016/j.ins.2009.12.010

Heidari, A. A., Abbaspour, R. A., and Chen, H. (2019). Efficient boosted grey wolf optimizers for global search and kernel extreme learning machine training. *Appl. Soft Comput.* 81:105521. doi: 10.1016/j.asoc.2019.105521

Heidari, A. A., and Pahlavani, P. (2017). An efficient modified grey wolf optimizer with Lévy flight for optimization tasks. *Appl. Soft Comput.* 60, 115–134. doi: 10.1016/j.asoc.2017.06.044

Hu, J., Chen, H., Heidari, A. A., Wang, M., Zhang, X., Chen, Y., et al. (2021). Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, feature selection. *Knowledge-Based Syst.* 213:106684. doi: 10.1016/j.knosys.2020.106684

Hu, Q.-X., Tian, J., and He, D.-J. (2017). Wheat leaf lesion color image segmentation with improved multichannel selection based on the Chan-Vese model. *Comp. Electr. Agri.* 135, 260–268. doi: 10.1016/j.compag.2017.01.016

Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., Rehman, M. H. U., and Javed, K. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Comp. Electr. Agri.* 153, 12–32. doi: 10.1016/j.compag.2018.07.032

Jia, H., Ma, J., and Song, W. (2019a). Multilevel thresholding segmentation for color image using modified moth-flame optimization. *IEEE Access.* 7, 44097–44134. doi: 10.1109/ACCESS.2019.2908718

Jia, H., Peng, X., Song, W., Lang, C., Xing, Z., and Sun, K. (2019b). Multiverse optimization algorithm based on lévy flight improvement for multithreshold color image segmentation. *IEEE Access.* 7, 32805–32844. doi: 10.1109/ACCESS.2019.2903345

Jothiaruna, N., and Sundar, K. J. A. (2020). "A segmentation method for comprehensive color feature with color-to-grayscale conversion using SVD and Region-Growing Method," in First *International Conference on Sustainable Technologies for Computational Intelligence*, Luhach AK, et al. Editors. (Thanjavur, India, Springer). doi: 10.1007/978-981-15-0029-9_24

Kotte, S., Rajesh Kumar, P., and Injeti, S. K. (2018). An efficient approach for optimal multilevel thresholding selection for gray scale images based on improved differential search algorithm. *Ain Shams Eng. J.* 9, 1043–1067. doi: 10.1016/j.asej.2016.06.007

Kumari, C. U., Prasad, S. J., Mounika, G., and Ieee, A. (2019). "Leaf disease detection: feature extraction with K-means clustering and classification with ANN," in *Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication*. (Erode, India, IEEE). doi: 10.1109/ICCMC.2019.8819750

Kuo, R. J., Lin, J. Y., and Nguyen, T. P. Q. (2020). An application of sine cosine algorithm-based fuzzy possibilistic c-ordered means algorithm to cluster analysis. *Soft Comput.* 25, 3469–3484. doi: 10.1007/s00500-020-05380-y

Liang, J. J., Qin, A. K., Suganthan, P. N., and Baskar, S. (2006). Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. *Ieee Transact. Evolut. Comput*. 10, 281–295. doi: 10.1109/TEVC.2005.857610

Liu, L., Zhao, D., Yu, F., Heidari, A. A., Li, C., Ouyang, J., et al. (2021a). Ant colony optimization with Cauchy and greedy Levy mutations for multilevel COVID 19 X-ray image segmentation. *Comput. Biol. Med*. 136:104609. doi: 10.1016/j.compbiomed.2021.104609

Liu, L., Zhao, D., Yu, F., Heidari, A. A., Ru, J., Chen, H., et al. (2021b). Performance optimization of differential evolution with slime mould algorithm for multilevel breast cancer image segmentation. *Comput. Biol. Med*. 138:104910. doi: 10.1016/j.compbiomed.2021.104910

Lv, M., Zhou, G., He, M., Chen, A., Zhang, W., and Hu, Y. (2020). Maize leaf disease identification based on feature enhancement and dms-robust alexnet. *IEEE Access* 8, 57952–57966. doi: 10.1109/ACCESS.2020.2982443

Merzban, M. H., and Elbayoumi, M. (2019). Efficient solution of Otsu multilevel image thresholding: A comparative study. *Expert Syst. Appl*. 116, 299–309. doi: 10.1016/j.eswa.2018.09.008

Mirjalili, S., Mirjalili, S. M., and Lewis, A. (2014). Grey wolf optimizer. *Adv. Eng. Softw*. 69, 46–61. doi: 10.1016/j.advengsoft.2013.12.007

Mittal, H., and Saraswat, M. (2018). An optimum multi-level image thresholding segmentation using non-local means 2D histogram and exponential Kbest gravitational search algorithm. *Eng. Appl. Artif. Intell*. 71, 226–235. doi: 10.1016/j.engappai.2018.03.001

Moayedi, H., and Mosavi, A. (2021a). Suggesting a stochastic fractal search paradigm in combination with artificial neural network for early prediction of cooling load in residential buildings. *Energies* 14:1649. doi: 10.3390/en14061649

Moayedi, H., and Mosavi, A. (2021b). An innovative metaheuristic strategy for solar energy management through a neural networks framework. *Energies* 14:1196. doi: 10.3390/en14041196

Moayedi, H., and Mosavi, A. (2021c). Electrical power prediction through a combination of multilayer perceptron with water cycle ant lion and satin bowerbird searching optimizers. *Sustainability* 13:2336. doi: 10.3390/su13042336

Moayedi, H., and Mosavi, A. (2021d). Synthesizing multi-layer perceptron network with ant lion biogeography-based dragonfly algorithm evolutionary strategy invasive weed and league champion optimization hybrid algorithms in predicting heating load in residential buildings. *Sustainability* 13:3198. doi: 10.3390/su13063198

Mousavi, S. M., Ghasemi, M., Dehghan Manshadi, M., and Mosavi, A. (2021). Deep learning for wave energy converter modeling using long short-term memory. *Mathematics* 9:871. doi: 10.3390/math9080871

Nenavath, H., and Jatoth, R. K. (2018). Hybridizing sine cosine algorithm with differential evolution for global optimization and object tracking. *Appl. Soft Comput*. 62, 1019–1043. doi: 10.1016/j.asoc.2017.09.039

Niu, P., Niu, S., and Chang, L. (2019). The defect of the Grey Wolf optimization algorithm and its verification method. *Knowledge-Based Syst*. 171, 37–43. doi: 10.1016/j.knosys.2019.01.018

Nosratabadi, S., Ardabili, S., Lakner, Z., Mako, C., and Mosavi, A. (2021). Prediction of food production using machine learning algorithms of multilayer perceptron and agriculture ANFIS. *Agriculture*. 11:408. doi: 10.3390/agriculture11050408

Oliva, D., Hinojosa, S., Elaziz, M. A., and Ortega-Sánchez, N. (2018). Context based image segmentation using antlion optimization and sine cosine algorithm. *Multimed. Tools Appl*. 77, 25761–25797. doi: 10.1007/s11042-018-5815-x

Qu, C. W., Zeng, Z. L., Dai, J., Yi, Z. J., and He, W. (2018). A modified sine-cosine algorithm based on neighborhood search and greedy levy mutation. *Comput. Intell. Neurosci*. 2018:4231647. doi: 10.1155/2018/4231647

Rodriguez-Esparza, E., Zanella-Calzada, L. A., Oliva, D., Heidari, A. A., Zaldivar, D., Pérez-Cisneros, M., et al. (2020). An efficient Harris hawks-inspired image segmentation method. *Expert Syst. Appl*. 155:113428. doi: 10.1016/j.eswa.2020.113428

Seifi, A., Ehteram, M., Singh, V. P., and Mosavi, A. (2020). Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and Sustainability ANN. *Sustainability* 12:4023. doi: 10.3390/su12104023

Setiadi, D. I. M. (2020). PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimed. Tools Appl*. 80, 8423–8444. doi: 10.1007/s11042-020-10035-z

Shaikh, R. P., Dhole, S., and Ieee, A. (2017). "Citrus leaf unhealthy region detection by using image processing technique," *2017 International Conference of Electronics, Communication and Aerospace Technology*. (Coimbatore, India, IEEE). doi: 10.1109/ICECA.2017.8203719

Shi, B., Ye, H., Zheng, L., Lyu, J., Chen, C., Heidari, A. A., et al. (2021). Evolutionary warning system for COVID-19 severity: Colony predation algorithm enhanced extreme learning machine. *Comput. Biol. Med*. 136:104698. doi: 10.1016/j.compbiomed.2021.104698

Song, S., Wang, P., Heidari, A. A., Wang, M., Zhao, X., Chen, H., et al. (2020). Dimension decided Harris hawks optimization with Gaussian mutation: Balance analysis and diversity patterns. *Knowledge-Based Syst*. 2020:106425. doi: 10.1016/j.knosys.2020.106425

Subramani, K., Periyasamy, S., and Theagarajan, P. (2019). Double line clustering based colour image segmentation technique for plant disease detection. *Curr. Med. Imaging* 15, 769–776. doi: 10.2174/1573405614666180322130242

Tang, H., Xu, Y., Lin, A., Heidari, A. A., Wang, M., Chen, H., et al. (2020). Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted k-nearest neighbor classifiers. *IEEE Access*. 8, 35546–35562. doi: 10.1109/ACCESS.2020.2973763

Tavoosi, J., Zhang, C., Mohammadzadeh, A., Mobayen, S., and Mosavi, A. H. (2021). Medical image interpolation using recurrent type-2 fuzzy neural network. *Front. Neuroinform*. 15:e667375. doi: 10.3389/fninf.2021.667375

Thaher, T., Heidari, A. A., Mafarja, M., Dong, J. S., and Mirjalili, S. (2020). "Binary Harris Hawks optimizer for high-dimensional, low sample size feature selection," in *Evolutionary Machine Learning Techniques* (Nablus, Palestine, Springer). doi: 10.1007/978-981-32-9990-0_12

Villalón, C. L. C., Stützle, T., and Dorigo, M. (2020). "Grey wolf, firefly and bat algorithms: three widespread algorithms that do not contain any novelty," in *International Conference on Swarm Intelligence* (Springer). doi: 10.1007/978-3-030-60376-2_10

Wang, A., Zhang, W., and Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Comp. Electr. Agri*. 158, 226–240. doi: 10.1016/j.compag.2019.02.005

Wang, J., He, J., Han, Y., Ouyang, C., and Li, D. (2013). An Adaptive Thresholding algorithm of field leaf image. *Comp. Electr. Agri*. 96, 23–39. doi: 10.1016/j.compag.2013.04.014

Wang, Z., Bovik, A. C., and Sheikh, H. R. (2004). Simoncelli EImage quality assessment: From error visibility to structural similarity. *Ieee Transact. Image Process*. 13, 600–612. doi: 10.1109/TIP.2003.819861

Wang, Z., Wang, K., Pan, S., and Han, Y. (2018). Segmentation of crop disease images with an improved K-means clustering algorithm. *Appl. Eng. Agric*. 34, 277–289. doi: 10.13031/aea.12205

Xiong, Y., Liang, L., Wang, L., She, J., and Wu, M. (2020). Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset. *Comp. Electr. Agri*. 177:105712. doi: 10.1016/j.compag.2020.105712

Yan, L., Zhang, J., and Ieee, A. (2018). "Image segmentation of rice blast disease based On two - dimensional histogram in HSI space," in *2018 13th World Congress on Intelligent Control and Automation*. (Changsha, China, IEEE). doi: 10.1109/WCICA.2018.8630477

Yang, F., Moayedi, H., and Mosavi, A. (2021). Predicting the degree of dissolved oxygen using three types of multi-layer perceptron-based artificial neural networks. *Sustainability* 13:9898. doi: 10.3390/su13179898

Zhang, L., Zhang, L., Mou, X. Q., and Zhang, D. (2011). FSIM: a feature similarity index for image quality assessment. *Ieee Transact. Image Process*. 20, 2378–2386. doi: 10.1109/TIP.2011.2109730

Zhang, Q., Wang, Z., Heidari, A. A., Gui, W., Shao, Q., Chen, H., et al. (2021). Gaussian Barebone Salp Swarm Algorithm with Stochastic Fractal Search for medical image segmentation: A COVID-19 case study. *Comp. Biol. Med*. 139:104941. doi: 10.1016/j.compbiomed.2021.104941

Zhang, S., You, Z., and Wu, X. (2019). Plant disease leaf image segmentation based on superpixel clustering and EM algorithm. *Neural Comput. Appl*. 31, 1225–1232. doi: 10.1007/s00521-017-3067-8

Zhao, D., Liu, L., Yu, F., Heidari, A. A., Wang, M., Liang, G., et al. (2020b). Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy. *Knowledge-Based Syst.* 2020:106510. doi: 10.1016/j.knosys.2020.106510

Zhao, D., Liu, L., Yu, F., Heidari, A. A., Wang, M., Oliva, D., et al. (2020a). Ant colony optimization with horizontal and vertical crossover search: fundamental visions for multi-threshold image segmentation. *Expert Syst. App.* 167:114122. doi: 10.1016/j.eswa.2020.114122

Zhao, S., Wang, P., Heidari, A. A., Chen, H., Turabieh, H., Mafarja, M., et al. (2021). Multilevel threshold image segmentation with diffusion association slime mould algorithm and renyi's entropy for chronic obstructive pulmonary disease. *Comp. Biol. Med.* 2021:104427. doi: 10.1016/j.compbiomed.2021.104427

Zhou, Y., Yang, X., Ling, Y., and Zhang, J. (2018). Meta-heuristic moth swarm algorithm for multilevel thresholding image segmentation. *Multimed. Tools Appl.* 77, 23699–23727. doi: 10.1007/s11042-018-5637-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 December 2021
doi: 10.3389/fpls.2021.753603

Ag-YOLO: A Real-Time Low-Cost Detector for Precise Spraying With Case Study of Palms

Zhenwang Qin[1], *Wensheng Wang*[1]*, *Karl-Heinz Dammer*[2], *Leifeng Guo*[1]* and *Zhen Cao*[1]

[1] *Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China,* [2] *Leibniz Institute for Agricultural Engineering and Bioeconomy, Department Engineering for Crop Production, Potsdam, Germany*

To date, unmanned aerial vehicles (UAVs), commonly known as drones, have been widely used in precision agriculture (PA) for crop monitoring and crop spraying, allowing farmers to increase the efficiency of the farming process, meanwhile reducing environmental impact. However, to spray pesticides effectively and safely to the trees in small fields or rugged environments, such as mountain areas, is still an open question. To bridge this gap, in this study, an onboard computer vision (CV) component for UAVs is developed. The system is low-cost, flexible, and energy-effective. It consists of two parts, the hardware part is an Intel Neural Compute Stick 2 (NCS2), and the software part is an object detection algorithm named the Ag-YOLO. The NCS2 is 18 grams in weight, 1.5 watts in energy consumption, and costs about $66. The proposed model Ag-YOLO is inspired by You Only Look Once (YOLO), trained and tested with aerial images of areca plantations, and shows high accuracy (F1 score = 0.9205) and high speed [36.5 frames per second (fps)] on the target hardware. Compared to YOLOv3-Tiny, Ag-YOLO is 2× faster while using 12× fewer parameters. Based on this study, crop monitoring and crop spraying can be synchronized into one process, so that smart and precise spraying can be performed.

Keywords: object detection, precise spraying, embedded AI, YOLO, NCS2

OPEN ACCESS

Edited by:
Dongjian He,
Northwest A&F University, China

Reviewed by:
Jun Liu,
Weifang University of Science and Technology, China
Baofeng Su,
Northwest A&F University, China

***Correspondence:**
Wensheng Wang
wangwensheng@caas.cn
Leifeng Guo
guoleifeng@caas.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *05 August 2021*
Accepted: *23 November 2021*
Published: *24 December 2021*

Citation:
Qin Z, Wang W, Dammer K-H, Guo L and Cao Z (2021) Ag-YOLO: A Real-Time Low-Cost Detector for Precise Spraying With Case Study of Palms. Front. Plant Sci. 12:753603. doi: 10.3389/fpls.2021.753603

1. INTRODUCTION

1.1. Motivation and Background

Areca catechu L. is also known as betel palm. It is cultivated mainly in tropical areas as South East Asia, India, South Pacific, and some African and Caribbean regions (Heatubun et al., 2012). The seed (areca nut) harvested is chewed in most cases because of the stimulating effect of its alkaloids. In a word, it is an importantly high-value crop. In the Hainan Island of China, this crop provides a livelihood to more than 2 million people in rural areas. Unfortunately, this cultivar has been suffering from the yellow leaf disease (YLD) that may lead to the decay and wilt of the palms. Luo et al. (2001) employed various methods to prove that the areca YLD in Hainan is an infectious disease caused by phytoplasma, including electron microscope observation, tetracycline antibiotic injection diagnosis, Polymerase Chain Reaction (PCR) technology. According to the studies, pests play a vital role in the spreading of viruses and phytoplasma. Therefore, spraying pesticides on the palms constantly is an effective way to prevent YLD from spreading.

With the advent of the Internet of Things (IoT), especially the rapid evolution of the UAV technology, combined with image data analytics, PA technologies have been developed, to increase productivity and at the same time reduce environmental impact. The PA technology focuses on about 20 relevant applications (Radoglou-Grammatikis et al., 2020), among which aerial monitoring and crop spraying are the most common. Spraying UAVs carry different types of equipment developed (Xiongkui et al., 2017; Lan and Chen, 2018; Yang et al., 2018) to spray pesticides to the crops in small fields or rugged environments, such as mountain areas. However, since an aerial crop monitoring process has not been synchronized with the spraying platform, today's UAVs spray the entire area uniformly with pesticides. Safe and effective spraying must be performed in areca protection due to the fact that: 1) In areca plantations, there is a certain distance from one palm to another; 2) Pesticides can have several side effects on the biotic and abiotic environment and bear a risk to harm human health (Horrigan et al., 2002); 3) Palms are irregular, especially in height, which increases the risk of UAV crashes.

For this purpose, this study develops an onboard CV component for spraying UAVs. The system takes RGB data acquired by the low-cost onboard camera as input, inferences and then sends instructions to the flight management unit (FMU) of the UAV in a real-time manner. As a result, pesticides can be applied on a per-plant basis, with a variable dosage subject to the severity of plant diseases. To be specific, the key feature of this system is to perform object detection in real-time.

In object detection tasks, deep learning (DL) methods significantly outperform other existing approaches due to their robustness to the diversity of targets. Nevertheless, the powerful performance of DL often comes with a high computation complexity and intensive memory demand, mainly required by the convolutional layers in convolutional neural network (CNN). For a high-end Graphics Processing Unit (GPU), this is not a problem. However, UAVs are tight constraints in computational power, memory size, and energy consumption. We solve the issue by extending the computing capacity with embedded hardware, then developing a new algorithm to fit. Embedded AI computing options are investigated, including graphics processing units, vision processing units, and field-programmable gate arrays. Today, commercial products are on the market, such as NVIDIA (https://www.nvidia.com/), Intel (https://www.intel.com), and MYiR (http://www.myir-tech.com/). As listed in **Table 1**, Intel NCS2 has both the least weight and power consumption. Besides, for a battery-powered device, those features are of great advantage.

1.2. Scope and Contribution

The overall goal of this study is to develop an object detection algorithm, which can run on NCS2 in real-time. We study the efficient object detection algorithms optimized for resources-constraint hardware and propose a novel model, as it is derived from the famous YOLO (Redmon et al., 2016), and used for agricultural purposes. Hence, we call it Ag-YOLO. Specifically, the summary of the contributions presented in this study is the following:

- We provide a thorough, complete description of the design, deployment, and assessment of an intelligent real-time agricultural object detection system based on embedded AI.
- By proposing the Ag-YOLO object detection algorithm and testing it on the NCS2, we demonstrate that a DL-based CV algorithm can be implemented on resource-constraint hardware, to deal with real-life PA challenges. On the most cost-efficient embedded AI device, the NCS2, our Ag-YOLO can achieve 36.5 fps with satisfying accuracy. The accuracy of Ag-YOLO is always higher than YOLOv3-Tiny in different input dimensions, and the highest accuracy of Ag-YOLO is 0.7655.
- We demonstrate a whole process to build an efficient object detector for palms. This method is easy to propagate to other cash crops such as pitaya, citrus.
- We developed a tool that is used for data training and transforming PC models to the NCS2 platform.
- We propose the "channel reorganization" block to adapt the ShuffleNet-v2 (Ma et al., 2018) backbone to NCS2, which shows the best speed performance.

1.3. Article Structure

The remainder of the study is organized as follows: Section 2 reviews related works on smart UAVs and Embedding AI. The proposed Ag-YOLO in this study is presented in Section 3. The experimental results, as well as a comparison to the baseline categorization and discussion, are presented in Section 4. Finally, Section 5 provides a summary of the study.

2. RELATED WORK

In this section, we review vision-based smart UAV applications. Furthermore, efficient embedded object detection algorithms are then discussed. Finally, works based on improved YOLO are presented.

2.1. Vision-Based Smart UAV Applications

In recent studies of UAVs, Intel NCS2 (https://www.intel.com/content/www/us/en/developer/tools/neural-compute-stick/overview.html), NVIDIA Jetson Nano (https://developer.nvidia.com/embedded/jetson-nano), NVIDIA Jetson TX2 (https://developer.nvidia.com/embedded/jetson-tx2) are used as companion computers in vision-based smart UAV applications to process aerial imagery. Then the output result is used to control the UAV's FMU. Dobrea and Dobrea (2020) places two embedded companion computers, a Raspberry Pi (RPi) and a Jetson Nano, on a HoverGames quadcopter to follow a pre-programmed flight route and simultaneously detect humans as well as of warning the system operator to reinforce the quarantine zones for epidemic prevention purposes. In the field of early fire detection and alarm, Nguyen et al. (2021) implements a real-time fire detection solution for vast area surveillance using the UAV with an integrated visual detection and alarm system. The system includes a low-cost camera, a lightweight companion

TABLE 1 | Main specifications of the candidate platforms in this study.

	Intel NCS2	nVidia Jetson Nano	MYiR ZU3EG
Features size	73 × 36 mm	70 × 45 mm	100 × 70 mm
HW accelerator	Myriad X VPU	128-core nVidia Maxcell GPU	Xilinx ultraScale
CPU	N./A.	Arm A57	MPSoC XCZU3EG (4-core Arm A53)
Peak performance	150 GFLOPs	472 GFLOPs	1.2 TFLOPs
Data precision	FP16	FP16/FP32	FP32
Nominal power	1.5 W	10 W	10 W
Weight	18 g*	140 g	150 g

**Weight of NCS2 is not including the outer shell.*

computer, a flight controller, and localization-and-telemetry modules. A Jetson Nano is used to support real-time detection, achieving a speed of 26 fps. In Afifi et al. (2019), the authors built a robust, real-time pedestrian detection system on Jetson TX2 for monitoring pedestrians by a UAV. Barisic et al. (2019) built a vision-based system for real-time detection and following of UAVs. The system achieves a real-time performance of 20 fps. Earlier, in Rabah et al. (2018), a small CPU RPi is used. In Alsalam et al. (2017), the authors developed an autonomous UAV using an Odroid U3+ and ROS to fulfill vision-based onboard decision making for remote sensing (RS) and PA.

2.2. Efficient Object Detection Algorithms

To detect objects in real-time with an embedded device, an efficient algorithm is required. DL-based object detection technology, which has rapidly developed since the mid-2000s, has overcome the limitations of the performance of other existing technologies, and their capabilities are similar to those of humans or sometimes exceed human abilities.

Among all the DL-based object detection frameworks, the YOLO-series (Redmon et al., 2016; Redmon and Farhadi, 2017; Farhadi and Redmon, 2018; Bochkovskiy et al., 2020) are widely used in various applications based on object detection in recent years due to their outstanding performance in terms of latency. In addition, the YOLO series algorithms also provide a trade-off between speed and accuracy, which allows researchers to apply them in different scenarios.

Although YOLOv4 (Bochkovskiy et al., 2020) has been released recently, YOLOv4 does not make any revolutionary improvement in architecture aspect to its forefather. YOLOv3 is still one of the most widely used detectors in the industry due to the limited computation resources and the insufficient software support in various practical applications. Hence, we choose YOLOv3 (Farhadi and Redmon, 2018) as our starting point and adopt some "Bag of freebies" strategies from YOLOv4. Specifically, the model we choose is the "lighter" version of YOLOv3, called Tiny-YOLOv3, which was designed with speed in mind and is generally reported as one of the better performing models in the aspect of speed and accuracy trade-off.

A YOLO-family detector is composed of backbone, neck, and head. The backbone is responsible for feature extraction, the neck synthesizes the features from backbone, and the head classifies the objects and labels the bounding boxes. As for the backbone part, there have been rising interests in improving it to achieve better speed in embedded devices, such as Howard et al. (2017), Ma et al. (2018), Wang et al. (2018), and Zhang et al. (2018). We investigate the effect by replacing them in YOLOv3-Tiny in the next section.

2.3. Improved Work Based on YOLO

Huang et al. (2018) proposed YOLO-Lite for bringing object detection to non-GPU computers. YOLO-Lite achieved 21 fps on a non-GPU computer and 10 fps after being implemented onto a website with only 7 layers and 482 million FLOPS. This speed is 3.8× faster than SSD Mobilenetv1, the fastest state-of-the-art model at that time. However, performances on embedding systems were not investigated.

Kim et al. (2020) investigated the performance degradation of spiking neural networks (SNNs) and presented the first spiked-based object detection model, called Spiking-YOLO. Spiking-YOLO achieves remarkable results that are comparable (up to 98%) to those of Tiny-YOLO on non-trivial datasets, PASCAL VOC, and MS COCO. Furthermore, Spiking-YOLO on a neuromorphic chip consumes approximately 280 times less energy than Tiny-YOLO and converges 2.3–4 times faster than previous SNN conversion methods.

Wong et al. (2019) introduced YOLO Nano, a highly compact deep CNN for embedded object detection designed using a human-machine collaborative design strategy, running on a Jetson AGX Xavier embedded module at different power budgets. At 15 and 30 W power budgets, YOLO Nano achieved inference speeds of ∼ 26.9 and ∼ 48.2 fps, respectively. The model size of YOLO Nano was 4.0 MB, which is 15.1× smaller than Tiny YOLOv2 and 8.3× smaller than Tiny YOLOv3. Despite being much smaller in model size, it achieved an mAP of 69.1% on the VOC 2007 test dataset, which is ∼ 12 and ∼ 10.7% higher than that of Tiny YOLOv2 and Tiny YOLOv3, respectively. Jetson AGX Xavier is a high-end embedded device, which is not considered in this study.

xYOLO is proposed in Barry et al. (2019) to detect balls and goal posts at ∼ 10 fps, on a piece of low-end hardware, the RPi 3 B, in a RoboCup Humanoid Soccer competition, compared to Tiny-YOLO which achieved 0.14 fps.

Hurtik et al. (2020) presented Poly-YOLO, which improves YOLOv3 in three aspects. It is more precise, faster, and able

to realize instance segmentation. Poly-YOLO has only 60% of parameters of the YOLOv3 but improves the accuracy by a relative 40%.

In Pham et al. (2020), Minh-Tan Pham et. al, designed YOLO-fine which is based on the state-of-the-art YOLOv3 with the main purpose of increasing the detection accuracy for small objects while being light and fast to enable real-time prediction within further operational contexts, providing the best compromise between detection accuracy (highest mAP), network size (smallest weight size), and prediction time (able to perform real-time prediction). No latency data are provided in this study.

3. MATERIALS AND METHODS

To the best of our knowledge, there is no public datasets for palms available, so we build one for this study.

3.1. Dataset

Four experimental sites in Sanya (18°15′10"N 109°30′42" E)[1] were selected to collect aerial images. In the experimental sites, areca palms of different ages were grown (from 2 years old to more than 20 years old). In addition, their spatial distance from palm to palm varies too. Some plantations were also heterogeneous in terms of the individual trunk volume. Therefore, a high object variation was guaranteed. Sanya is a city on the Hainan Island of China. The images were collected by using a DJI Phantom 4,[2] of which the camera resolution was 5,472 × 3,078, and the aperture was F 1/2.8. This UAV hovered above the palms at a height ranging from 2 to 10 m to take photos so that images could be in different scales. The angle of view of the camera was between 45 and 90°. An example image is shown in **Figure 1** (For better illustration, this image was taken from a much higher altitude for readers who have no idea of what an areca plantation is like). More than 1,000 aerial images were taken at different times on August 2, August 3, October 5, and November 4, 2018. The collecting time ranges from early morning, midday to sunset to get different sunlight conditions. Among them, 400 images were picked and then resized to a smaller dimension (the maximum length was set to 1,500 to speed up the processing time), which were labeled with an open-source software named labelImg.[3] After that, 300 images were randomly selected and saved into the "Training" dataset and the rest 100 images were saved into the "Testing" dataset.

It is worth noting, there are two reasons we did not use all 1,000 images. First, it took a lot of manually repetitive labor to label images; Second, with "Data Augmentation" and "Hard Negative Mining" technologies, for a small model which only detects objects of one class, 400 images are enough.

During training, we used the "Data Augmentation" method to expand the training dataset. A labeled image was cropped randomly and resized to the dimension of the network (e.g., 416 × 416). Then, it was transformed into the HSV (Hue, Saturation, and Value) color space, so that any of the H, S, and V values could be adjusted randomly for the simulation of illumination changes, or color changes. For example, reducing the H value by 1 or 2 to simulate a little less illumination intensity. After that, the image was transformed back to RGB (Red, Greed, and Blue) color space. This was implemented by using the tool *RQNet*, and the default parameter values for "Data Augmentation" are used.

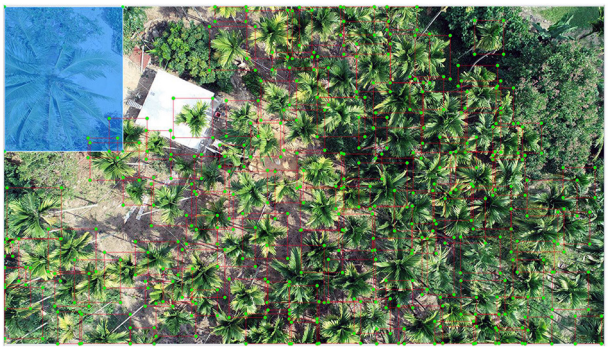


FIGURE 1 | Areca palm plantation at the experimental site labeled with red quadrates. The blue semitransparent box covers a coconut palm (part of the background).

3.2. Data Training

All frameworks were trained on an end-to-end basis in a single T1060 GPU optimized by Adam (Kingma and Ba, 2014) at the initial learning rate of 0.001. Each mini-batch has 10 images. Therefore, one epoch includes 15 mini-batches. This study resized the input dimension to (352, 352), (384, 384), (416, 416), (448, 448), (480, 480), (512, 512), (544, 544), and (576, 576) for every epoch randomly.

For every model, the parameters were initialized by the Xavier method. After using strong data augmentation, we found that ImageNet pre-training is no more beneficial, we, thus, train all the following models from scratch. By adopting Leaky ReLU as an activation function and using Gaussian distribution initialized parameter, all the models were easily converged in hundreds of thousands of iterations, taking 2 ∼ 4 days on an ASUS TUF Gaming FX86FM laptop. The value of Gaussian parameters $\mu = 0$, $\sigma = (16n)^{-0.5}$ where n refers to the number of weight elements.

3.2.1. L2-Norm Regularization and NCS2 Deployment

Regularization has been introduced into DL for a long time, which brings in additional information for the prevention of over-fitting. The L2-norm regularization can be expressed as follows:

$$L = \sum_{(x,y)} l(f(x, W), y) + \lambda \sum_{w} w^2 \quad (1)$$

where λ refers to the super parameter, and in YOLO articles, it is referred to as "decay" or "weight decay," and set to 0.0005. The x and y denote the coordinate value of the feature map, and w denotes the parameters of the model.

[1]https://en.wikipedia.org/wiki/Sanya
[2]https://www.dji.com/cn/phantom-4-rtk?site=brandsite&from=nav
[3]https://github.com/tzutalin/labelImg or https://pypi.org/project/labelImg/

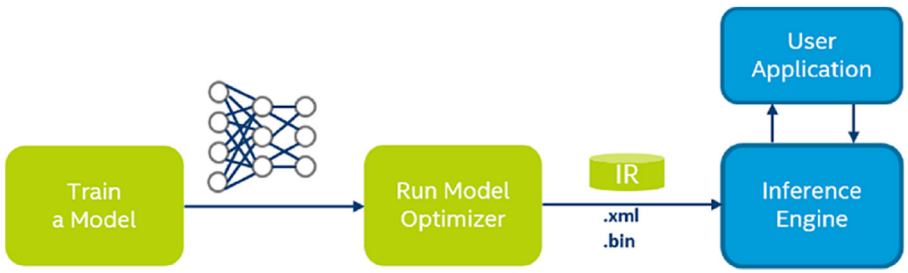


FIGURE 2 | Network training and deployment on Neural Compute Stick 2 (NCS2).

In this study, the model was trained on a PC, and tested and used on NCS2, as described in **Figure 2**. The well-trained model on a PC is composed of a large number of parameters in FP32 format. They need to be parsed and converted to FP16 format and organized in a form that the NCS2 driver can understand.

Two pieces of C++ tools have been developed for data training and model testing on NCS2, the *RQNet*, and the *OpenVINOMyriad* (refer to https://github.com/rossqin/). The *RQNet* was used for data training and model evaluation under the Windows operating system, which was also used to convert models used by the *OpenVINOMyriad* to run on NCS2. During the training phase, the CUDA 11.1 and cuDNN 8.0 libraries were used, and all the parameters were in standard 32-bit float point values (FP32). However, NCS2 only supports 16-bit "half precision" float point values (FP16), which can express values within the range ±65,504 with the minimum value above 1 being 1 + 1/1024. To minimize the accuracy loss while the parameters are being quantified from FP32 to FP16, the parameters should be small enough. However, in the case that λ starts with a small value, a model with a bunch of huge value parameters beyond FP16 might be obtained, especially in the first layer. To avoid this case, λ was set to 0.01 during the first 100 k iterations and then set to 0.001.

3.2.2. Loss

A YOLOv3 (Farhadi and Redmon, 2018) object detector predicts bounding boxes using dimension clusters as prior boxes. For each bounding box, there are 4 corresponding predicted values, i.e., t_x, t_y, t_w, and t_h. When the center of the object is in the cell offset from the top left corner of the image by (c_x, c_y), and the prior box has the dimension o (p_w, p_h), then the prediction values correspond to

$$b_x = \sigma(t_x) + c_x \tag{2a}$$
$$b_y = \sigma(t_y) + c_y \tag{2b}$$
$$b_w = p_w \cdot e^{t_w} \tag{2c}$$
$$b_h = p_h \cdot e^{t_h} \tag{2d}$$

As in **Figure 3**:

For YOLOv3, a prediction loss comprises 3 parts, i.e., the object loss L_{obj}, the classification loss: L_{cls}, and the coordinate loss L_{box}.

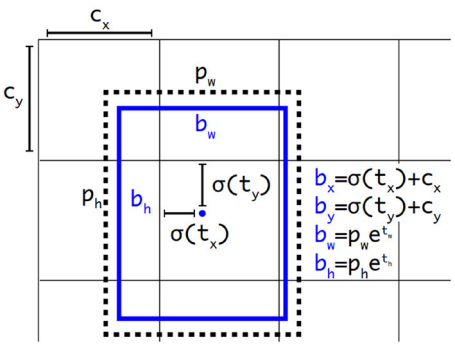


FIGURE 3 | Bounding boxes with dimension priors and location prediction for YOLOv3.

$$Loss = L_{obj} + L_{cls} + L_{box} \tag{3}$$

Where

$$L_{obj} = \lambda_{noobj} \sum_{i}^{S^2} \sum_{j}^{B} \mathbb{1}_{i,j}^{noobj} (c_i - \hat{c}_i)^2 + \lambda_{obj} \sum_{i}^{S^2} \sum_{j}^{B} \mathbb{1}_{i,j}^{obj} (c_i - \hat{c}_i)^2 \tag{4a}$$

$$L_{cls} = \lambda_{cls} \sum_{i}^{S^2} \sum_{j}^{B} \mathbb{1}_{i,j}^{obj} \sum_{c \in classes} p_i(c) log\left(\hat{p}_i(c)\right) \tag{4b}$$

$$L_{box} = \lambda_{box} \sum_{i}^{S^2} \sum_{j}^{B} \mathbb{1}_{i,j}^{obj} \left(2 - w_i \times h_i\right) \tag{4c}$$

$$\times \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2\right] \tag{4d}$$

where S denotes the size of the feature map to be predicted, B represents the prior boxes count, $\mathbb{1}_{i,j}^{obj}$ refers to the fact that the i-th cell and the j-th prior box are responsible for one ground truth, and $\mathbb{1}_{i,j}^{noobj}$ refers to the opposite.

In this study, only palms need to be detected, therefore, it is always assumed that $L_{cls} = 0$. Moreover, Focal Loss (Lin et al.,

2017) is used in L_{obj} to increase the recall rate R and suppress the erroneous recall rate FP :

$$L_{obj} = -\lambda_{obj} \sum_{i}^{S^2} \sum_{j}^{B} \mathbb{1}_{i,j}^{obj} \alpha \, (1 - c_i)^{\gamma} \log(c_i) \tag{5}$$

where the parameters of α and γ were set to 0.5 and 0.2, respectively. Focal Loss accelerates convergence during the training process because it gives much higher penalty weights to poor predictions. If a prediction value is close to the ground truth value, then it's much less important to keep minimizing the gap. As a result, the training process can pay more attention to poor predictions.

In terms of L_{box}, CIOU Loss proposed in Zheng et al. (2020) is used just as Bochkovskiy et al. (2020) do, express as follows,

$$L_{box} = \lambda_{box} \sum_{i}^{S^2} \sum_{j}^{B} \mathbb{1}_{i,j}^{obj} \left(1 - \mathtt{IoU}_i + \frac{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}{c_i^2} + \frac{v_i^2}{(1 - \mathtt{IoU}_i) + v_i}\right) \tag{6}$$

Where c_i^2 is the area of the minimum box containing the prediction box and ground truth box.

$$v_i = \frac{4}{\pi} \left(\arctan \frac{\hat{w}_i}{\hat{h}_i} - \arctan \frac{w_i}{h_i}\right) \tag{7}$$

The values of λ_{obj} and λ_{box} are set to 1 and 0.2, respectively, but when the model is hard to converge, λ_{box} can be adjusted according to the condition.

3.2.3. Network Slimming

Studies demonstrated that accuracy can be improved by increasing the layers (deeper layers) (Simonyan and Zisserman, 2014) or the channels in layers (wider layers) (Howard et al., 2017). In this study, an initial network architecture wider and deeper enough was used, and the network was made to learn its structural sparsity. Besides, network slimming was used as well, which was introduced in the previous section. The slimming was performed on a well-trained network when the importance of the γ parameters in the BN layers was further learned. In addition, no regularization was imposed on the parameters in the convolutional layers, and the model was re-trained after pruning.

The training scheme in network slimming was similar to that of normal training, specifically, λ (the weight_decay value) started from 0.01 and then 0.001 after 100 k iterations.

3.2.4. Background Training

As shown in **Figure 1**, in most areca plantations, the contrast between foreground and background was not very obvious. Almost all the images were green with the variation from light to dark green, except for some yellowish-brown spots which were the YLD diseased palm individuals.

All the palms in the dataset were labeled, and to increase the accuracy, the predictor was trained so that the background will not be predicted as objects. Those prior boxes not overlapped with any of the ground truths were defined as "background boxes." During the training phase, in the case that a background box was predicted as an object, in another word, the confidence value was larger than the threshold value (e.g., 0.5), the predictor was punished. This extra work strengthened the predictor's ability to distinguish objects from the background, and decreased false-positive predictions.

3.3. Prior Boxes

Some studies involved the in-depth investigation of prior box selection in the YOLO model. It is empirically believed that some losses of accuracy were originated from the unequal distribution of the ground truth by anchors, in another word, one specific prior box in a cell response predicted more than the ground truth, so that during the training process, there is no way to learn all the ground truth. One solution to this problem is to avoid this conflict. For example, to use better designed prior boxes array or a bigger prior box collection. It is unnecessary to use more predictors for a light model (e.g., in Mazzia et al., 2020) if the backbone network has enough representational power, which is because that more predictors bring more computation complexity. In the study, the k-means was used to pick prior boxes for our model over the dataset, for example, in YOLOv3-Tiny, at first, $k = 6$ was set to get a box array of (23,23), (35,36), (48,49), (64,66), (90,91), and (147,157), which was referred to as "def-anchors" in the later section, with all the images normalized to (416,416). Since the smallest box is bigger than a high-resolution cell grid (16,16) in both width and height, another box array of (10,14), (27,23), (37,58), (75,64), (93,104), and (187,163) was used, which was referred to as "cust-anchors" to see what happens if there is one smaller prior box than the smaller cell grid. This study also used k = 8 to get another box array of (19,19), (27,29), (37,36), (43,48), (58,57), (71,75), (99,101), (158,169), referred to as "8-anchors."

3.4. The Structure of YOLOv3-Tiny and Some Related Components

The basic YOLOv3-Tiny architecture is shown in **Figure 4**. As the one-stage object detector, it comprises a backbone network, one or more prediction heads, and corresponding necks. The backbone network of YOLOv3-Tiny is named *Darknet18*, which is framed by a red box. The YOLOv3-Tiny has 2 prediction heads in different scales and corresponding necks, which synthesizes and organizes high-level features of the input images.

The YOLOv3-Tiny uses an intuitive neck structure, which takes quite a lot of computation overhead. The ResBlock component proposed in PeleeNet (Wang et al., 2018), as in **Figure 5**, was used to compare with the original neck.

Some efficient backbone networks were evaluated, and it was found that the backbone networks proposed by MobileNet v2 and those proposed by ShuffleNet v2 showed time performance. In this study, they were also adapted for better

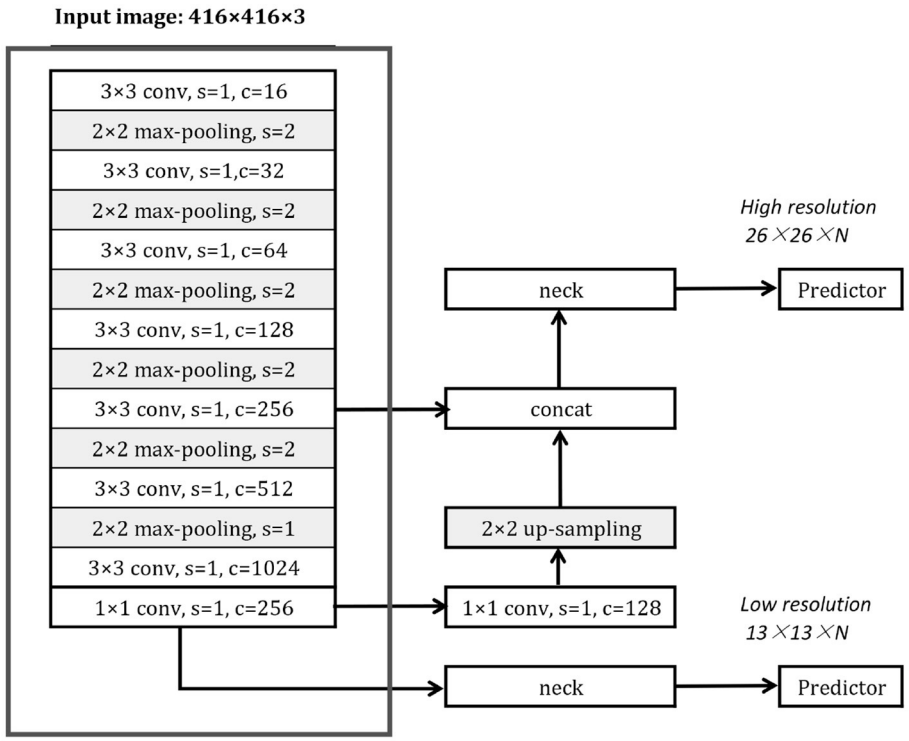


FIGURE 4 | The YOLOv3-Tiny framework, s: stride, c: channel.

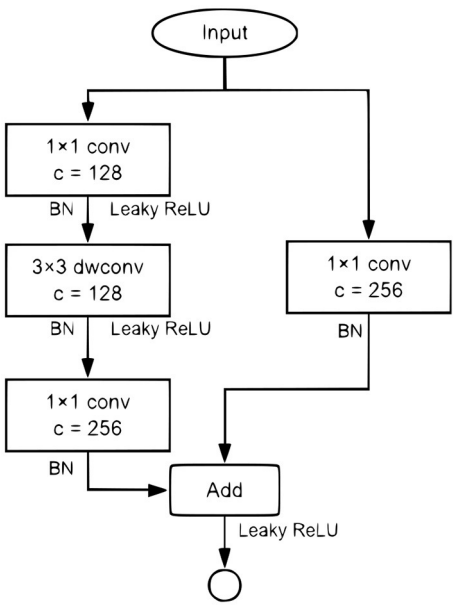


FIGURE 5 | The ResBlock in PeeleNet.

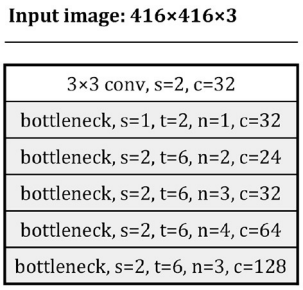


FIGURE 6 | The compact MobileNet v2 backbone.

performance. **Figure 6** shows a compact version of MobileNet v2's backbone, and the bottleneck module is as defined in the original study.

ShuffleNet v2 building blocks were used to build our backbone except for some small modifications. There were two reasons, the first is that the *Channel Shuffle* operation (as shown in **Figure 7A**) was not supported by the NCS2 hardware, therefore, *Channel Reorganization* was used to achieve the same or similar effect, (as shown in **Figure 7B**); Second, in this study, we found that in MobileNet v2 (as shown in **Figures 8A,C**), an activation layer followed a 3 × 3 depth-wise convolutional layer instead of a 1 × 1 one, and it showed higher accuracy (as shown in **Figures 8B,D**). This study compared the three cases as follows, 1) A 3 × 3 depth-wise convolutional layer followed by an activation layer without activation layer for the 1 × 1 convolutional layer; 2) No activation layer for 3 × 3 depth-wise convolutional layers, and a 1 × 1

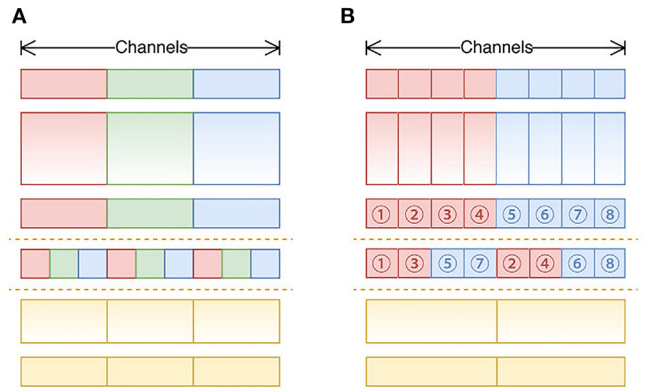


FIGURE 7 | (A) The standard channel shuffle operation in ShuffleNet; **(B)** channel reorganization operation in our study.

convolutional layer is followed by an activation layer; 3) Both convolutional layers are followed by activation layers, with the details described in Section 4.

Figure 9 shows our backbone architecture.

3.5. Evaluation Metrics

The inference time on NCS2, Parameters Size (of the model), Billion FLoating Operations (BFLOPs), F1 score, and Intersection of Union (IoU) were applied to evaluate the detection performance. Inference time denotes the duration of detecting objects from one image which is a speed metric, BFLOPs which is a computational complexity metric, F1 score combines the performance evaluation of the recall and the precision of the detection, therefore, it is widely applied as the evaluation index in many previous studies of object detection when there is only one object category. The expression of the precision, recall, and F1 is expressed as follow:

$$P = \frac{TP}{TP + FP} \tag{8a}$$

$$R = \frac{TP}{TP + FN} \tag{8b}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{8c}$$

where P denotes the precision, R refers to the recall, TP represents the true positives, FP is the false positives, and FN denotes the false negatives. The definition of IoU is shown in **Figure 10**, which measures the intersection area of the predicted object boundary box and the ground truth, thereby evaluating the location accuracy of the predicted boundary box of the prediction.

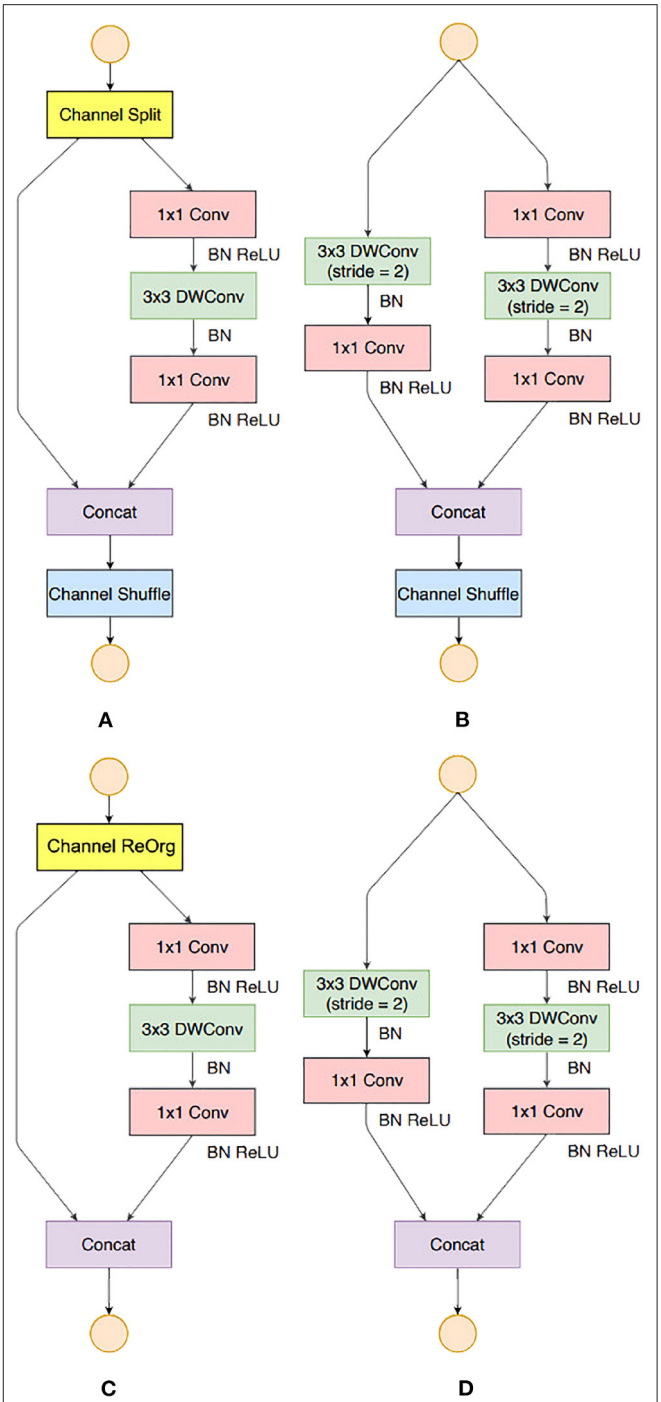


FIGURE 8 | Building Blocks of ShuffleNet v2 and this study. **(A)** the basic ShuffleNet v2 unit; **(B)** the ShuffleNet v2 unit for spatial down sampling (2×); **(C)** our basic unit; **(D)** our unit for spatial down sampling (2×).

4. RESULTS

In this section, a performance comparison was made among the vanilla YOLOv3-Tiny, the YOLOv3-Tiny with training improvement skills, the algorithm comprised of the Darknet18 backbone and the ResBlock neck, the algorithm comprised of efficient backbone and the ResBlock neck.

Based on the resulted data, we obtained Ag-YOLO.

4.1. Improving YOLOv3-Tiny Without Change to Structure

The Darknet18 backbone is very simple, which can be deemed as the simplest structure according to the principle proposed by VGG16. It is also a very good structure to inspect the performance data for NCS2. In this section, the impact of different prior boxes and background training was investigated.

Table 2 shows the accuracy incremental improvement obtained by background training and prior boxes selection.

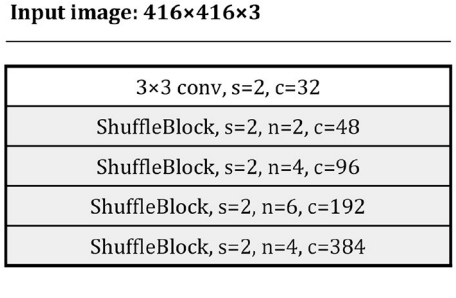


FIGURE 9 | The ShuffleNet v2 backbone. Each line describes a sequence of one or more identical (modulo stride) layers, repeating for n times. All layers in the same sequence have the same number c of output channels. The first layer of each sequence has a stride s and all others use stride 1. All spatial convolutions use 3 × 3 kernels.

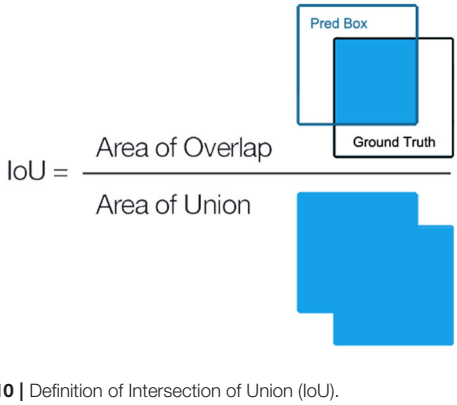


FIGURE 10 | Definition of Intersection of Union (IoU).

By training background and using more prior boxes, the accuracy was improved with a very small computational overhead, which was not reflected in the inference time on NCS2. The reason is that the hand-selected prior boxes failed to bring notable accuracy changes due to the fact that there were not enough small objects in the trained or tested images. What has to be pointed out is that the "Inference Time on NCS2" refers to the net computation time spent on the device, excluding the image decoding and data transfer.

4.2. Using the ResBlock Neck

The ResBlock component is proposed in PeleeNet (Wang et al., 2018). As shown in **Figure 5**.

Table 3 shows the performance improvement brought by the ResBlock neck. In this comparison, 8 prior boxes are used in predictors (4 by each).

The F1 score in both $\mathbf{IoU_{0.5}}$ and $\mathbf{IoU_{0.75}}$ markedly improved at a cost of a slight drop in average IoU, which was acceptable. Moreover, the parameters were reduced from 8.668 million to 6.848 million, and BFLOPs were reduced from 5.449 to 4.109 billion, leading to an inference time shrink of more than 10 ms, i.e., about 27% of the original value.

4.3. Network Slimming

In the darknet source code,[4] L2-norm regularization was imposed on the weights with $\lambda = 0.0005$. This value was too small, which leads to big parameters beyond FP16. Starting from $\lambda = 0.01$, this study found that in some channels, all the parameters tended to be zero, therefore, those channels were removed to reduce computation. However, by imposing L1-norm regularization on the γ parameters in batch normalization layers, a better result was achieved. This study pruned all the channels with $|\gamma| < 0.5$, and the pruned results were shown in **Table 4**.

Network slimming failed to bring notable performance improvement in terms of inference time, on the contrary, it induced a little degradation to IoU and F1 scores in this experiment. However, in some architectures, it generated a smaller model.

4.4. Using Efficient Backbones

4.4.1. SqueezeNet

Table 5 shows the performance variation while Darknet18 was replaced by SqueezeNet. SqueezeNet achieved better

[4]https://pjreddie.com/darknet/

TABLE 2 | Accuracy incremental improvement.

Model	Average IoU	F1 Score		Parameters Size	BFLOPs	Inference Time on NCS2
		$\mathbf{IoU_{0.5}}$	$\mathbf{IoU_{0.75}}$			
Default-anchors-no-bg	0.8186	0.9236	0.7358	8.634 M	5.436	37.75 ms
Default-anchors-bg	0.8202	0.9250	0.7503	8.634 M	5.436	37.75 ms
Custom-anchors-bg	0.8170	0.9278	0.7422	8.634 M	5.436	37.75 ms
8-anchors-bg	**0.8266**	**0.9300**	**0.7685**	8.668 M	5.449	37.75 ms

Bold values indicate the best performance.

TABLE 3 | Accuracy improvement by ResBlock.

Model	Average IoU	F1 Score		Parameters Size	BFLOPs	Inference Time on NCS2
		$IoU_{0.5}$	$IoU_{0.75}$			
YOLOv3-Tiny built-in	**0.8266**	0.9300	0.7685	8.668 M	5.449	37.75 ms
ResBlock	0.8247	**0.9432**	**0.7777**	**6.848 M**	**4.109**	**27.42 ms**

Bold values indicate the best performance.

TABLE 4 | Performance data by pruning.

Model	Average IoU	F1 Score		Parameters Size	BFLOPs	Inference Time on NCS2
		$IoU_{0.5}$	$IoU_{0.75}$			
ResBlock	0.8247	0.9432	0.7777	6.848 M	4.109	27.42 ms
Pruned	0.8237	0.9433	0.7625	6.548 M	3.949	27.33 ms

TABLE 5 | Performance data of SqueezeNet.

Model	Average IoU	F1 Score		Parameters Size	BFLOPs	Inference Time on NCS2
		$IoU_{0.5}$	$IoU_{0.75}$			
Default-anchors-no-bg	0.8186	0.9236	0.7358	8.634 M	5.436	37.75 ms
SqueezeNet	0.8346	0.9304	0.7860	1.186 M	5.176	56.67 ms

TABLE 6 | Performance data of MobileNet v2.

Model	Average IoU	F1 Score		Parameters Size	BFLOPs	Inference Time on NCS2
		$IoU_{0.5}$	$IoU_{0.75}$			
Default-anchors-no-bg	0.8186	0.9236	0.7358	8.634 M	5.436	37.75 ms
MobileNet v2	0.8443	0.9592	0.8240	1.082 M	2.095	65.32 ms

TABLE 7 | Performance data of ShuffleNet v2.

Model	Average IoU	F1 Score		Parameters Size	BFLOPs	Inference Time on NCS2
		$IoU_{0.5}$	$IoU_{0.75}$			
Default-anchors-no-bg	0.8186	0.9236	0.7358	8.634 M	5.436	37.75 ms
ShuffleNet v2(1)	**0.8349**	0.9448	**0.7893**	813 K	1.033	26.23 ms
ShuffleNet v2(2)	0.8278	**0.9513**	0.7668	**711 K**	**0.985**	**25.96 ms**
ShuffleNet v2(3)	0.8178	0.9404	0.7394	878 K	1.071	27.60 ms

Bold values indicate the best performance.

performance, e.g., better IoU and fewer parameters. However, it doubled computation complexity and increased the inference time.

Table 6 shows its performance. With the bottleneck microstructure, improvements were made in terms of IoU, F1 score, and BFLOP. However, the inference time was much shorter (about 2.5 times that by Darknet18).

Table 7 shows the results. The second one has the best performance, and for the third one, more non-linearity leads to worse performance. Combined with a modified version of ShuffleNet-v2 backbone, a ResBlock neck, and a YOLOv3 head, a new YOLO framework were proposed. We temporarily named it *Ag-YOLO* because it could be used for agricultural purposes.

4.5. Models Tested on NCS2

All models were converted to OpenVINO-version and tested on an NCS2 device. The host was a Windows 10 laptop and the data was transferred *via* USB3 protocol also supported by the RPi 4 computer. Due to the data precision loss, performance degradation occurred for all models. As seen from **Table 8**, the Ag-YOLO improved the original YOLOv3-Tiny version significantly in terms of both speed (about 6 frames more in a second) and accuracy (about 0.2 increase in F1 score). The model

TABLE 8 | Models tested on NCS2.

Model	Features				Backbone	fps	F1 Score	IoU
	CIoU Loss	BG	ResBlock	Pruned				
1 (YOLOv3-Tiny)					Darknet18	20.7	0.9160	0.6959
2	✓				Darknet18	20.7	0.9276	0.7148
3	✓	✓			Darknet18	20.7	0.9302	0.7253
4	✓	✓	✓		Darknet18	26.2	0.9223	0.7209
5	✓	✓	✓	✓	Darknet18	26.3	0.9209	0.7108
6	✓	✓	✓	✓	PeleeNet	14.6	0.9211	0.7352
7	✓	✓	✓	✓	Compact MobileNet v2	13.0	0.9364	0.7410
8 (Ag-YOLO)	✓	✓	✓	✓	ShuffleNet v2 derived	**26.9**	0.9361	0.7395

IoU = 0.5, input dimensions: 416 × 416, confidence threshold = 0.4, and non-maximum-supress threshold = 0.5. Bold value indicates the best performance.

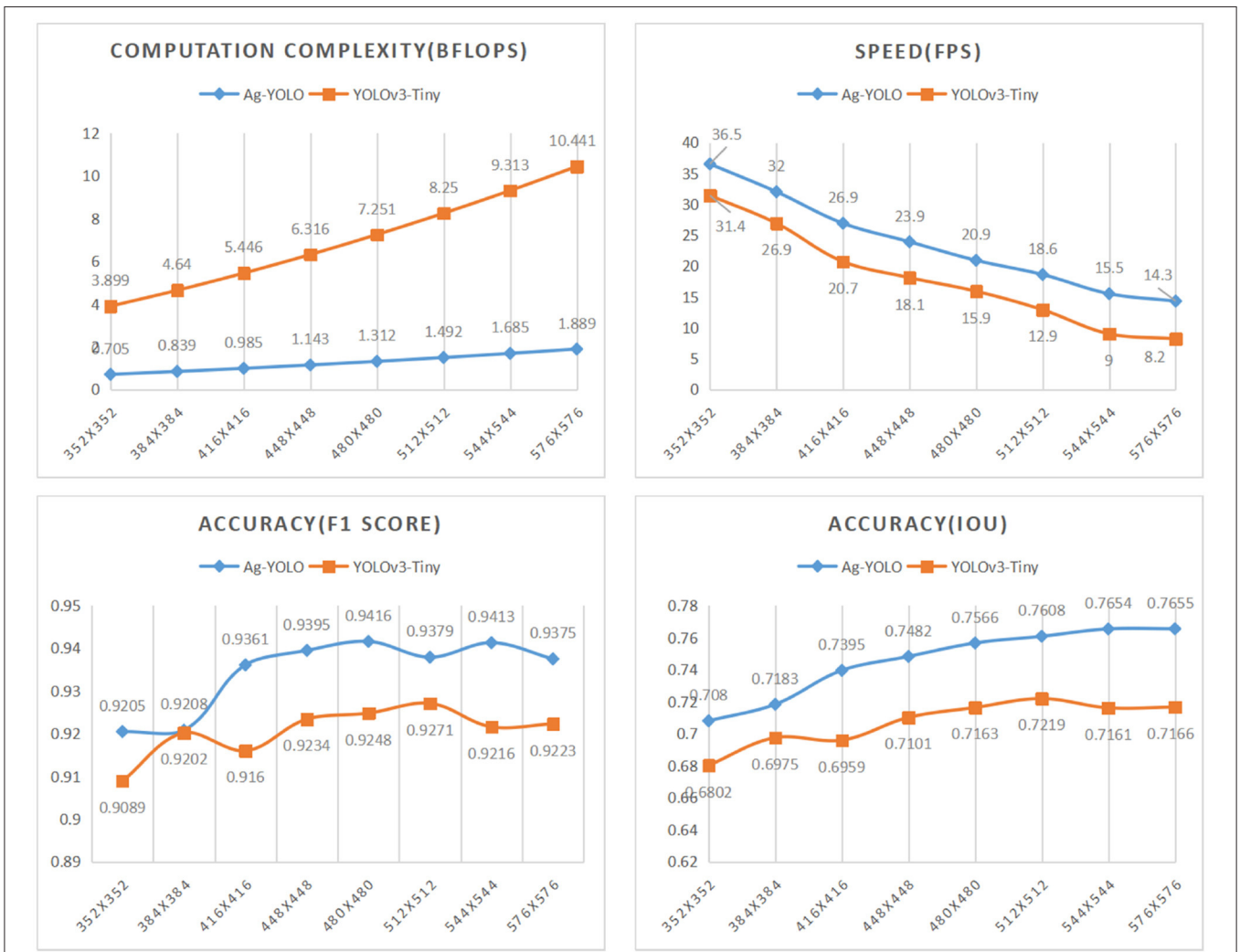


FIGURE 11 | Ag-YOLO vs. YOLOv3-Tiny under different input dimensions.

using a compact MobileNet v2 backbone surpassed our model a little in terms of F1 score and IoU, however, it takes double times to run.

When the input dimension was set to 352 × 352, Ag-YOLO achieved the speed of 36.5 fps with an F1 score of 0.9205 and IoU of 0.708 on NCS2, while YOLOv3-Tiny achieved similar

FIGURE 12 | Ag-YOLO run on NCS2 (Input dimension: 416 × 416). Images were taken by a UAV, different colors of the predicted square imply different confidence values, blue is low, and red is high. **(A)** R:92%, P:100%, F1:0.9583, IoU:0.7199. **(B)** R:92.86%, P:96.30%, F1:0.9455, IoU:0.7345.

accuracy at the speed of 18.1 fps with an input dimension of 448 × 448. Based on these data, Ag-YOLO is two times faster than YOLOv3-Tiny.

Different input dimensions: Performance of a model is also affected by the input dimension. As in the training phase, the input dimensions had been changed to 352 × 352, 384 × 384, 416 × 416, 448 × 448, 480 × 480, 512 × 512, 544 × 544, 576 × 576. **Figure 11** shows performance trends of Ag-YOLO and YOLOv3-Tiny. **Figure 12** is an example of the NCS2 output.

5. DISCUSSION

Abdulridha et al. (2019) applied a hyperspectral camera for the detection of citrus canker disease in citrus plantations. Modica et al. (2020) used UAV multi-spectral imagery to monitor the vigor in heterogeneous citrus and olive orchards. Ye et al. (2020) identified Fusarium wilt in bananas using supervised classification algorithms with UAV-based multi-spectral imagery. Those camera systems are characterized by expensiveness, difficulty in operation, relatively large size, and susceptibility to crash situations compared with RGB cameras.

The developed software, presented in this study, is specially adapted for use in embedded RGB-camera systems. With the increasing availability of UAVs that can spray pesticides, the algorithm can contribute to performing selective spraying. Therefore, the pesticides could be saved, thereby reducing the environmental impact and the economical costs of the farmer. In particular, cheap technology is necessary for the wide use of target-orientated selective spraying. Additionally, a cheap RGB-camera-controlled UAV spraying should also be affordable for small farmers.

The source code of this study is available at https://github.com/rossqin/RQNet, which can be used as a reference for the beginning researchers to develop their real-life AI applications instead of pursuing higher performance with new algorithms and the ever-increasing demand for higher computational power and memory requirements. In a specific agricultural CV task, for instance, object detection, the object category is usually one or few, therefore, it is possible to use a small and efficient DNN-based model to achieve a good result. This is proved in this study by exploring the YOLOv3-Tiny architecture and replacing the neck and backbone with different state-of-art efficient DNNs, such as SqueezeNet, MobileNet v2, and ShuffleNet v2. This study also uses network slimming to compress the models to obtain smaller models. In the meantime, this study trained all the models on a laptop and tested them on a low-cost hardware accelerator, i.e, the Intel NCS2. Our architecture, the Ag-YOLO, is comprised of a ShuffleNet v2-derived backbone, a ResBlock neck, and a YOLOv3 head, with only 813k parameters and 1.033 billion FLOPs, which is only 9.4 and 19% of the Darknet18 version, respectively. However, it brings better accuracy and inference time performance on the resource-constraint hardware NCS2, achieving a speed of 36.5 fps. Because a camera usually takes video at the frame rate of 24 fps, this is a REAL-TIME object detector. On the other hand, a compact version of the MobileNet v2 backbone leads to a better accuracy performance, although it takes more than twice BFLOPs and inference time. In a UAV or UGV auto-pilot use case, the host usually moves quite slow, and there is no need to process each frame from the onboard camera. For the tasks that emphasize accuracy, the compact version of the MobileNet v2 backbone presents a better option for Ag-YOLO. To obtain better accuracy, redundant information between successive frames can be utilized, as shown in Bozek et al. (2018).

6. CONCLUSION

This study decomposes the "YOLOv3-Tiny" into a backbone network, one or more necks, and corresponding heads. The backbone network extracts the features from an image, the necks synthesize the features that backbone network outputs, and the heads decode the information as required. This work improves YOLOv3-Tiny by replacing a more efficient backbone and better neck, in addition, we adopt some "Freebies" and "Back-of-Specials" such as CIOU Loss and more prior boxes in heads. Our work demonstrates that, a DNN-based CV algorithm can be implemented on resource-constraint device to deal with real-life PA challenge, even with the most costefficient embedded AI device, e.g., the NCS2. In addition, our Ag-YOLO can achieve 36.5 fps with satisfying accuracy. The accuracy of Ag-YOLO is always higher than that of YOLOv3-Tiny in different input dimensions, and the highest accuracy of Ag-YOLO reaches 0.7655. This experiment also demonstrated that a MobileNetv2-derived backbone showed better representational power, and a ShuffleNetv2-like backbone runs faster at the cost of a little accuracy degradation. Besides, both of them are superior in terms of computation intensity and memory usage. With this work, including the open-source toolset, it should be very easy to make their legacy agricultural machinery intelligent by using an onboard camera and an edge computing device.

7. FUTURE WORK

The proposed model Ag-YOLO has been proved to be competent in extending a UAV with little overhead, from cost to energy. The CV will be integrated into a practical spraying UAV *via* the Mavlink protocol to deal with the challenges in areca protection.

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 in the article/supplementary material.

AUTHOR CONTRIBUTIONS

WW: conceptualization. ZQ: methodology, software, writing—original draft preparation, data curation, and formal analysis. K-HD, ZQ, LG, and ZC: writing—review and editing. WW and LG: project administration. LG: funding acquisition. ZC: proofread. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was funded by National Natural Science Foundation of China (Grant No. 31860180), Major Science and Technology Program of Inner Mongolia Autonomous Region (Grant No. ZD20190039), Natural Science Foundation of Xinjiang Uygur Autonomous Region (Grant No. 2018D01A20), and Science and Technology Project of Inner Mongolia Autonomous Region (Grant No. 2021GG0341).

REFERENCES

Abdulridha, J., Batuman, O., and Ampatzidis, Y. (2019). Uav-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sens.* 11:1373. doi: 10.3390/rs11111373

Afifi, M., Ali, Y., Amer, K., Shaker, M., and ElHelw, M. (2019). Robust real-time pedestrian detection in aerial imagery on jetson tx2. *arXiv preprint* arXiv:1905.06653.

Alsalam, B. H. Y., Morton, K., Campbell, D., and Gonzalez, F. (2017). "Autonomous uav with vision based on-board decision making for remote sensing and precision agriculture," in *2017 IEEE Aerospace Conference* (Big Sky, MT: IEEE),1–12.

Barisic, A., Car, M., and Bogdan, S. (2019). "Vision-based system for a real-time detection and following of uav," in *2019 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS)* (Cranfield: IEEE), 156–159.

Barry, D., Shah, M., Keijsers, M., Khan, H., and Hopman, B. (2019). "Xyolo: a model for real-time object detection in humanoid soccer on low-end hardware," in *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)* (Dunedin: IEEE), 1–6.

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: optimal speed and accuracy of object detection. *arXiv preprint* arXiv:2004.10934.

Bozek, K., Hebert, L., Mikheyev, A. S., and Stephens, G. J. (2018). "Towards dense object tracking in a 2d honeybee hive," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 4185–4193.

Dobrea, D.-M., and Dobrea, M.-C. (2020). An autonomous uav system for video monitoring of the quarantine zones. *Roman J. Inf. Sci. Technol.* 23:S53. Available online at: https://www.romjist.ro/full-texts/paper654.pdf

Farhadi, A., and Redmon, J. (2018). *Yolov3: An incremental improvement. In Computer Vision and Pattern Recognition*. 1804–2767. Available online at: https://tethys.pnnl.gov/sites/default/files/publications/Redmonetal.pdf

Heatubun, C. D., Dransfield, J., Flynn, T., Tjitrosoedirdjo, S. S., Mogea, J. P., and Baker, W. J. (2012). A monograph of the betel nut palms (areca: Arecaceae) of east malesia. *Bot. J. Linnean Soc.* 168, 147–173. doi: 10.1111/j.1095-8339.2011.01199.x

Horrigan, L., Lawrence, R. S., and Walker, P. (2002). How sustainable agriculture can address the environmental and human health harms of industrial agriculture. *Environ. Health Perspect.* 110, 445–456. doi: 10.1289/ehp.02110445

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, R., Pedoeem, J., and Chen, C. (2018). "Yolo-lite: a real-time object detection algorithm optimized for non-gpu computers," in *2018 IEEE International Conference on Big Data (Big Data)* (Seattle, WA: IEEE), 2503-2510.

Hurtik, P., Molek, V., Hula, J., Vajgl, M., Vlasanek, P., and Nejezchleba, T. (2020). Poly-yolo: higher speed, more precise detection and instance segmentation for yolov3. *arXiv preprint* arXiv:2005.13243.

Kim, S., Park, S., Na, B., and Yoon, S. (2020). Spiking-yolo: spiking neural network for energy-efficient object detection. *Proc. AAAI Conf. Artif. Intell.* 34, 11270–11277. doi: 10.1609/aaai.v34i07.6787

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint* arXiv:1412.6980.

Lan, Y., and Chen, S. (2018). Current status and trends of plant protection UAV and its spraying technology in China. *Int. J. Precis. Agri. Aviat.* 1, 135–140. doi: 10.33440/j.ijpaa.20180101.0002

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: IEEE), 2980–2988.

Luo, D. Q., Chen, M. R., Ye, S. B., and Cai, X. Z. (2001). Study on pathogen identification of areca nut yellowing disease in hainan. *J. Trop. Crops* 22, 43–46. Available online at: https://en.cnki.com.cn/Article_en/CJFDTotal-RDZX200102007.htm

Ma, N., Zhang, X., Zheng, H.-T., 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.

Mazzia, V., Khaliq, A., Salvetti, F., and Chiaberge, M. (2020). Real-time apple detection system using embedded systems with hardware accelerators: an edge ai application. *IEEE Access* 8, 9102–9114. doi: 10.1109/ACCESS.2020.2964608

Modica, G., Messina, G., De Luca, G., Fiozzo, V., and Praticò, S. (2020). Monitoring the vegetation vigor in heterogeneous citrus and olive orchards. a multiscale object-based approach to extract trees' crowns from uav multispectral imagery. *Comput. Electron. Agric.* 175:105500. doi: 10.1016/j.compag.2020.105500

Nguyen, A., Nguyen, H., Tran, V., Pham, H. X., and Pestana, J. (2021). "A visual real-time fire detection using single shot multibox detector for uav-based fire surveillance," in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)* (Phu Quoc Island: IEEE), 338–343.

Pham, M.-T., Courtrai, L., Friguet, C., Lefèvre, S., and Baussard, A. (2020). Yolo-fine: one-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sens.* 12, 2501. doi: 10.3390/rs12152501

Rabah, M., Rohan, A., Talha, M., Nam, K.-H., and Kim, S. H. (2018). Autonomous vision-based target detection and safe landing for uav. *Int. J. Control Autom. Syst.* 16, 3013–3025. doi: 10.1007/s12555-018-0017-x

Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., and Moscholios, I. (2020). A compilation of uav applications for precision agriculture. *Comput. Netw.* 172:107148. doi: 10.1016/j.comnet.2020.107148

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* (Las Vegas, NV: IEEE), 779–788.

Redmon, J., and Farhadi, A. (2017). "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI: IEEE), 7263–7271.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556.

Wang, R. J., Li, X., and Ling, C. X. (2018). Pelee: A real-time object detection system on mobile devices. *arXiv preprint* arXiv:1804.06882.

Wong, A., Famuori, M., Shafiee, M., Li, F., Chwyl, B., and Chung, J. (2019). Yolo nano: A highly compact you only look once convolutional neural network for object detection. arxiv 2019. *arXiv preprint* arXiv:1910.01271. doi: 10.1109/EMC2-NIPS53020.2019.00013

Xiongkui, H., Bonds, J., Herbst, A., and Langenakens, J. (2017). Recent development of unmanned aerial vehicle for plant protection in east asia. *Int. J. Agric. Biol. Eng.* 10, 18–30. Available online at: http://www.ijabe.org/index.php/ijabe/article/view/3248

Yang, S., Yang, X., and Mo, J. (2018). The application of unmanned aircraft systems to plant protection in china. *Precision Agric.* 19, 278–292. doi: 10.1007/s11119-017-9516-7

Ye, H., Huang, W., Huang, S., Cui, B., Dong, Y., Guo, A., et al. (2020). Identification of banana fusarium wilt using supervised classification algorithms with uav-based multi-spectral imagery. *Int. J. Agric. Biol. Eng.* 13, 136–142. doi: 10.25165/j.ijabe.20201303.5524

Zhang, X., Zhou, X., Lin, M., and Sun, J. (2018). "Shufflenet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 6848–6856.

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D. (2020). Distance-iou loss: faster and better learning for bounding box regression. *Proc. AAAI Conf. Artif. Intel.* 34, 12993–13000. doi: 10.1609/aaai.v34i07.6999

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fpls.2021.810214

Minimalizing Non-point Source Pollution Using a Cooperative Ion-Selective Electrode System for Estimating Nitrate Nitrogen in Soil

Rui Su[1,2,3], Junfeng Wu[1], Jiandong Hu[1,2,3], Liuzheng Ma[1,2,3], Shakeel Ahmed[1,2,3], Yanyan Zhang[1,2,3], Mukhtar Iderawumi Abdulraheem[1,2,3,4], Zephania Birech[5], Linze Li[1,2,3], Can Li[1,2,3] and Wentao Wei[1,2,3]*

[1] Department of Electrical Engineering, Henan Agricultural University, Zhengzhou, China, [2] Henan International Joint Laboratory of Laser Technology in Agriculture Sciences, Zhengzhou, China, [3] State Key Laboratory of Wheat and Maize Crop Science, Zhengzhou, China, [4] Department of Agricultural Science Education, Oyo State College of Education, Lanlate, Nigeria, [5] Department of Physics, University of Nairobi, Nairobi, Kenya

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Ashok Kumar,
VIT University, India
Daming Dong,
Beijing Research Center for Information Technology in Agriculture, China
Babatunde Ewulo,
Federal University of Technology Akure, Nigeria

***Correspondence:**
Jiandong Hu
jdhu@henau.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *06 November 2021*
Accepted: *10 December 2021*
Published: *12 January 2022*

Citation:
Su R, Wu J, Hu J, Ma L, Ahmed S, Zhang Y, Abdulraheem MI, Birech Z, Li L, Li C and Wei W (2022) Minimalizing Non-point Source Pollution Using a Cooperative Ion-Selective Electrode System for Estimating Nitrate Nitrogen in Soil. Front. Plant Sci. 12:810214. doi: 10.3389/fpls.2021.810214

Nitrate nitrogen (NO_3^- -N) in the soil is one of the important nutrients for growing crops. During the period of precipitation or irrigation, an excessive NO_3^- -N readily causes its leaching or runoff from the soil surface to rivers due to inaccurate fertilization and water management, leading to non-point source pollution. In general, the measurement of the NO_3^--N relies upon the laboratory-based absorbance, which is often time-consuming, therefore not suitable for the rapid measurements in the field directly. Ion-selective electrodes (ISEs) support the possibility of NO_3^--N measurement by measuring the nitrate (NO_3^-) ions in soil quickly and accurately due to the high water solubility and mobility of NO_3^- ions. However, such a method suffers from a complicated calibration to remove the influences caused by both temperature and other ions in the measured solution, thus limiting field use. In this study, a kind of all-solid ISE system combined with a temperature sensor and a pH electrode is proposed to automatically measure the concentrations of the NO_3^--N. In this study, a soil water content calibration function was established, which significantly reduces a relative error (RE) by 13.09%. The experimental results showed that the stabilization time of this electrode system was less than 15 s with a slope of −51.63 mV/decade in the linear range of 10^{-5}–$10^{-2.2}$ mol/L. Both the limit of detection of 0.5 ppm of the NO_3^--N and a relative SD of less than 3% were obtained together with the recovery rate of 90–110%. Compared with the UV-Vis spectroscopy method, a correlation coefficient (R^2) of 0.9952 was obtained. The performances of this all-solid ISE system are satisfied for measuring the NO_3^--N in the field.

Keywords: nitrate nitrogen, all-solid nitrate ion-selective electrodes, water content calibration function, non-point source pollution, recovery rate

HIGHLIGHTS

- A cooperative ISEs system is proposed to measure soil nitrate nitrogen.
- The detection limit, RSD and the recovery rate meet practical level.
- Guide inaccurate fertilization and irrigation resulting in non-point source pollution.
- A water content calibration function is applied to improve the measurement precision.
- This ISEs system is validated with samples from the planting zones in central China.

INTRODUCTION

Nitrate nitrogen (NO_3^--N) is an important inorganic nutrient in soil that crops can absorb and use directly (Gebbers and Adamchuk, 2010). The fundamental criterion for determining the nitrogen (N) fertilizer utilization rate during crop growth is confirmed by its concentration in the soil (Burton et al., 2018; Nyameasem et al., 2020). Increasing N fertilizer dosage and application is usually one of the essential ways to boost crop productivity (Van Groenigen et al., 2015; Zhang and Yao, 2017). Both N fertilizer and animal dung contain a significant amount of NO_3^--N (Pennino et al., 2017). However, an excess of NO_3^--N is leached from the soil surface and then transported to rivers, lakes, and groundwater during precipitation or irrigation (Sadler et al., 2016; Velusamy et al., 2021). Thus, the improper management of N fertilizer leads to a large amount of NO_3^--N leaching and causes non-point source pollution (Cui et al., 2020; Mahmud et al., 2020). NO_3^--N leaching leads to the loss of crop nutrients (Sun et al., 2020). In contrast, it also damages both human and animal health, showing colorectal cancer and Non-Hodgkin lymphoma due to the excessive NO_3^--N concentration in drinking water (Szpak, 2014; Mary et al., 2018). Therefore, it is important to normalize the measurement of NO_3^--N in soil (Cho et al., 2018; Richa et al., 2021). Considering the concentrations of NO_3^--N in soil is readily influenced by environmental factors and changes rapidly, it is necessary to perform *in situ* measurement to evaluate the non-point source pollution (Wang et al., 2015; Ma et al., 2019b).

There are many traditional methods used to measure soil NO_3^--N, but they are not convenient for use in the field. For example, the UV-Vis spectrophotometry is currently used in laboratories to quantify NO_3^--N (Burton et al., 2020). Also, the method is professional in operation, and the related chemical reagents bring secondary pollution (Rogovska et al., 2018). Differently, the electrochemical methods exhibit fastness and effectiveness on measuring the concentrations of NO_3^--N in the soil, which is expected to achieve the *in situ* measurement (Ma et al., 2019a). Adamchuk et al. (2005) created a direct soil testing system that used ISEs to analyze soil nutrients and tracked the spatial variability in nutrient distribution, although the R^2 of NO_3^--N is only 0.41–0.51 in comparison with laboratory testing. Kim et al. (2007) built a sensor array consisting of N, phosphorus, and potassium-selective electrodes and an argentic/argentic chloride (Ag/AgCl) electrode as reference for the detection of NO_3^--N in soil with the R^2 value of only 0.89. Moreover, ISEs have also been applied to construct a soil nutrient mapping device, which can generally be mounted on agricultural machinery to map soil conditions over the farm terrain (Sibley et al., 2008). In further measurement of NO_3^--N, the R^2 value was reached to 0.96 by an improved ISE (Tully and Weil, 2014). Although the abovementioned methods have achieved great progress on NO_3^--N measurement, there are still many challenges associated with environmental factors and such cannot be used in the field. Temperature, salinity, and soil water content influence should be eliminated to improve precision results. In this study, an automatic cooperative measurement system for detecting the NO_3^--N in soil was developed. Specifically, this system consists of an all-solid-state ISE, a temperature electrode, and a pH electrode to accurately measure the NO_3^--N. Moreover, the developed cooperative system is expected to realize the *in situ* measurement of soil NO_3^--N and providing scientific soil fertilization recommendations at the grassroots level, thus reducing the non-point source pollution. In this system, a peristaltic pump was utilized to control this system working automatically through the microcontroller unit. To lessen the impacts from soil water content, a correction function was introduced to increase the accuracy of on-site measurements toward soil NO_3^--N. The signals from the cooperative ISE system are transferred to an upper computer in real time.

MATERIALS AND METHODS

Study Area and Soil Sampling

Field experiments were carried out in the wheat-maize rotation zones, including Xuchang City (113°47′08″E, 34°09′35″N), Henan Agricultural University, China (**Figure 1a**), and the rice growth field in Qianjiang City (112°39′07″E, 30°18′23″N), Hubei Province (**Figure 1b**). In this experiment, the field area of 300 m^2 was totally divided into four plots. The different fertilization modes were arranged in each plot of 50.0 m^2 (2.5 m × 20 m). Each plot has six rows of planted crops and a 1-m ridge erected between them. **Figure 1c** depicts the zones for soil sampling, including (1) control (no fertilization); (2) normal fertilization, i.e., the compound fertilizer of 50 kg/66.7 m^2 (the urea containing 46% N, calcium superphosphate containing 16% phosphorus pentoxide, and potassium sulfate containing 50% potassium oxide, at a ratio of 14:6:9), (3) under-fertilization (0.5 times), and (4) overfertilization (1.5 times). Other field management practices were identical in each plot. Five soil samples at the depth of 0–20 cm were collected from each plot by the plum blossom pattern method. Notably, 21 soil samples were randomly chosen from two zones, and each soil sample was repeatedly measured six times.

Cooperative Ion-Selective Electrode System

All-solid-state ISEs (cooperative ISEs) with small volume and fast responses were used to determine the ion concentration (activity) of NO_3^--N from the membrane potentials in the

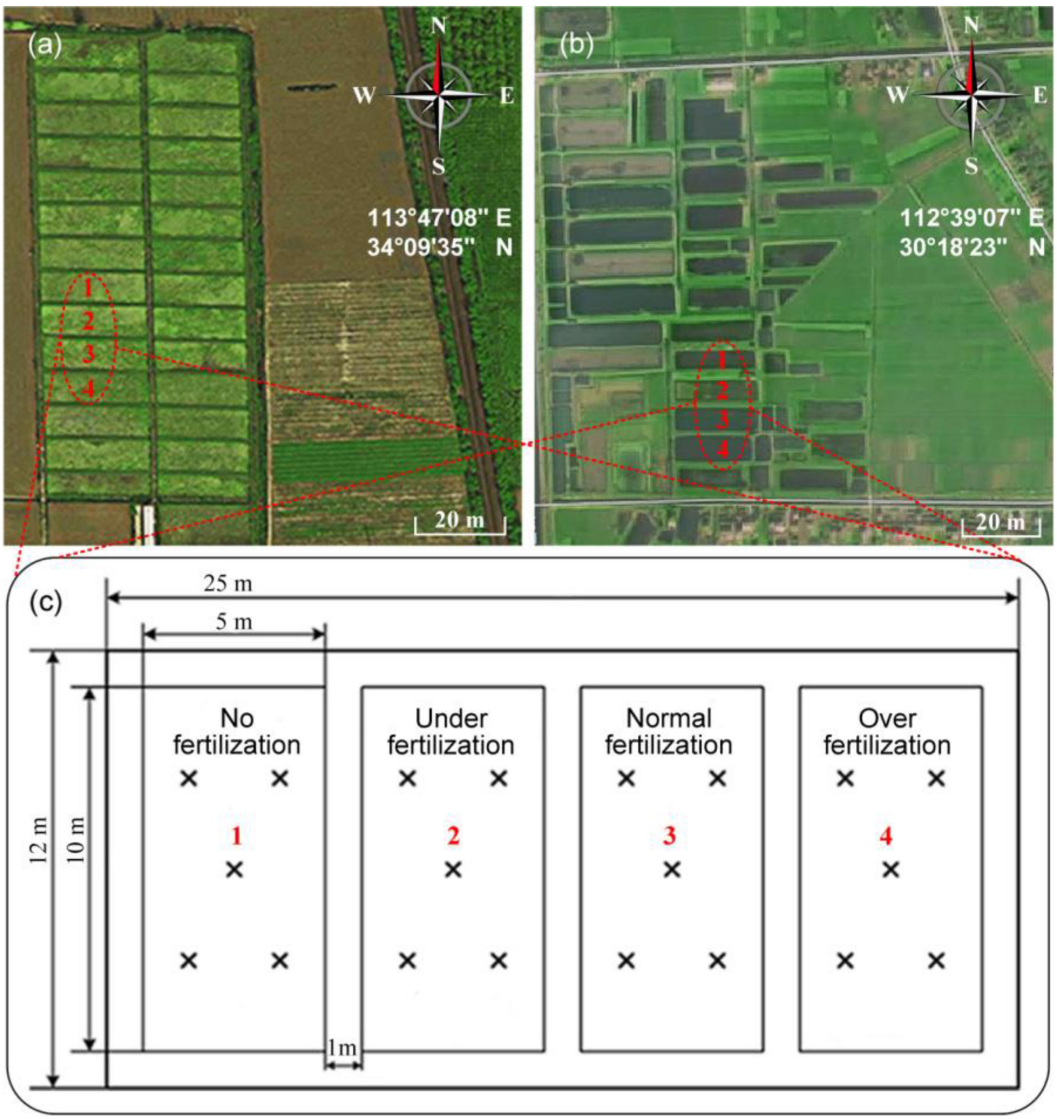


FIGURE 1 | Field experiment. The selected experiment place in **(a)** Xuchang (113°47′08″E, 34°09′35″N) and **(b)** Qianjiang (112°39′07″E, 30°18′23″N). **(c)** Field fertilization modes in the experiment place and "×" indicates the soil sampling locations.

solution. The dissolved NO_3^- ions diffuse through the polyvinyl chloride membrane in the ISEs when the electrode was immersed into the solution to test the NO_3^- ions. When the concentrations on both sides of the membrane reach equilibrium, the membrane potential difference becomes steady. Then, the logarithm of the NO_3^- concentrations in the external solution is proportional to measured electrode potential, as Eq. 1,

$$E = E^{\Theta} - \frac{2.303RT}{nF}\log\left[NO_3^-\right] \quad (1)$$

Where E is the electrode potential, E^{Θ} is standard apparent electrode potential, including potential difference of membrane/external solution, membrane/internal solution potential, and special interface/internal solution, R is a constant called the universal gas constant (8.314 J K^{-1} mol^{-1}), T is the absolute temperature (K), F is the Faraday constant (96,485 C/mol), and $n = 1$ is the charge transfer number for the reduction in NO_3^-.

Figure 2 depicts the cooperative ISE system for the determination of the NO_3^- -N concentrations. The measurement procedure involved: first, placing a soil sample in the pretreatment unit followed by pressing the software button on the LCD touch screen or using the APP in a phone to start the measurement. The soil weight is measured automatically and stored in the microcontroller unit. The peristaltic pump then draws both the deionized water and 2% (V/V) ionic strength adjustment buffer of 2M ammonium sulfate to the sample cell. In the sample cell, a microcontroller-controlled stepper motor alters the height of the water content sensor. Stirring of the sample solution is then achieved *via* an agitator for 3 min. Thereafter,

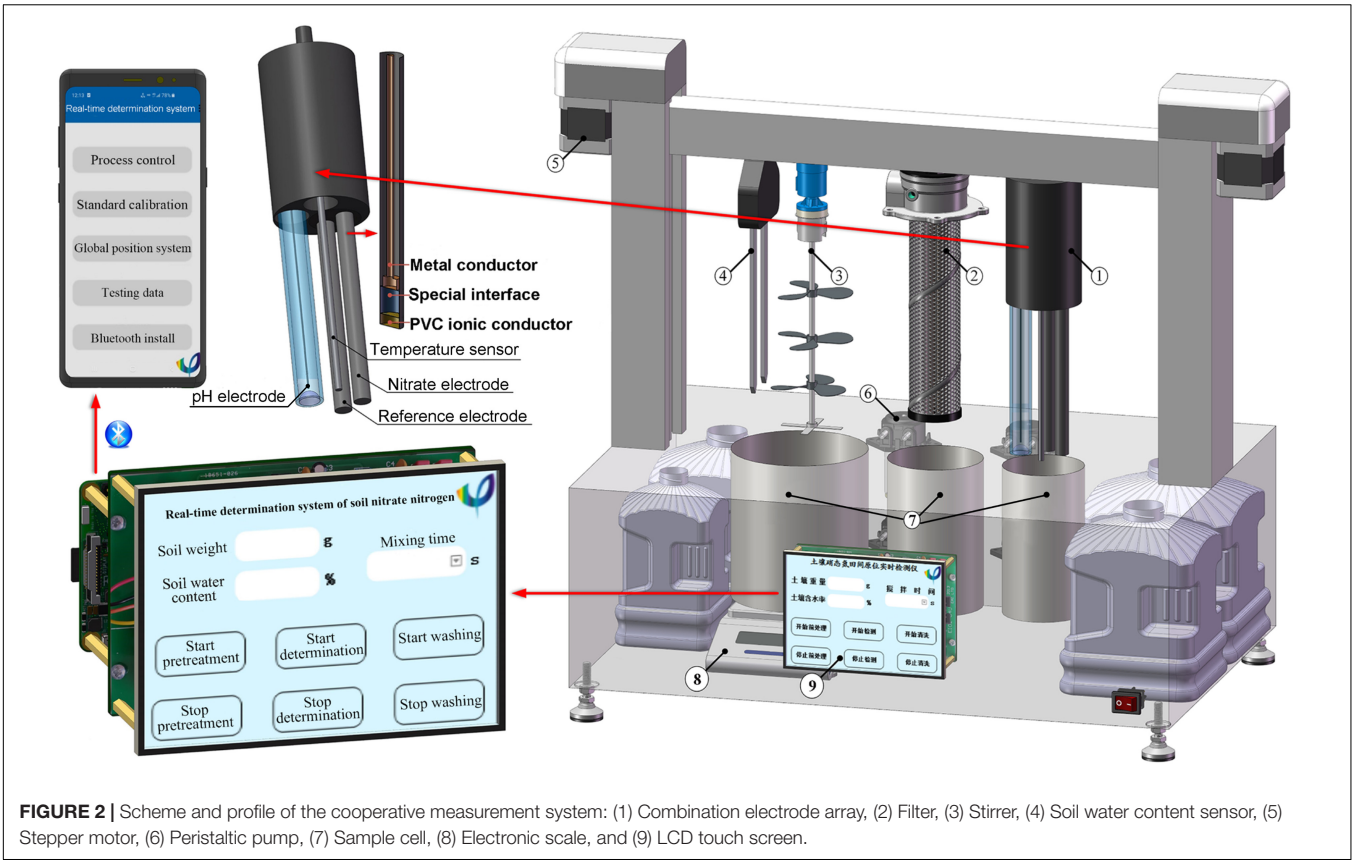


FIGURE 2 | Scheme and profile of the cooperative measurement system: (1) Combination electrode array, (2) Filter, (3) Stirrer, (4) Soil water content sensor, (5) Stepper motor, (6) Peristaltic pump, (7) Sample cell, (8) Electronic scale, and (9) LCD touch screen.

the peristaltic pump draws the supernatant to the detecting unit. In the measurement unit, the ISE system consisting of an NO_3^--N electrode, a pH electrode, and a temperature sensor then detects the soil NO_3^--N concentrations, pH, and temperature, respectively. The signal processing circuit transfers the potential difference generated by the electrode array to the microprocessor, and then, the concentration of NO_3^--N, which is calculated after the compensation, has been done with soil water content. The collected data are then stored in the microcontroller unit and transmitted to the APP in a phone through the Raspberry Pi's Bluetooth interface. Once the measurement is completed, the peristaltic pump launches again to automatically clean these electrodes with deionized water.

Ab initio Calculation

The geometries of NO_2^-, NO_3^-, and SO_4^{2-} were optimized at the M06-2X/6-311+G(2d,p) level of theory. The frequency calculation was performed at the same level of theory to ensure that the optimized configurations are located at the minimum of the potential surface. According to the benchmark of Zhao and Truhlar (2007), the UV-Vis spectrophotometer was calculated at the M06-2X/6-311+G(2d,p) level of theory with the assistance of time-dependent density functional theory (Li et al., 2020), and the solvation model density (Marenich et al., 2009) was used to represent the water solution situation. All the above *ab initio* calculations were performed by ORCA 5.0.1 programmer (Neese, 2011). The UV-Vis spectrophotometer (Feller Instrument limited business, Nanjing) was used to check the absorption peak positions of each interfering ion. The site of the maximum absorption peak of NO_3^- was obviously identified to be at a wavelength of 203 nm (**Figure 3A**). Thus, the standard curves for the UV-Vis spectra method can be obtained (**Figure 3B**). To confirm the maximum absorption peak of NO_3^- in UV-Vis absorbance spectra, the *ab initio* calculation is accurately performed first. Later, we can manage these experiments to establish the standard curve at the maximum absorption peak of NO_3^- The value of the maximum absorption peak seriously influences the correlation coefficient (R^2) obtained from both the ISE system and UV-Vis absorbance spectra.

RESULTS AND DISCUSSION

Water Content Calibration Function

Normally, the soil water contents in the field change from 5 to 20%. The soil with a water content of less than 5% is too dry for crop growth, and NO_3^--N is easily leached out away when the water content exceeds 25%. Notably, 35 measured samples with various water content (2, 5, 10, 15, 20, 25, and 30%) were obtained by adding deionized water into the five dried soil samples (marked as S_1, S_2, S_3, S_4, and S_5). The influence of soil water content on the NO_3^--N concentration measured from this ISE system is shown in **Figure 4**. Obviously, the influence of soil water content on ISE determination results cannot be

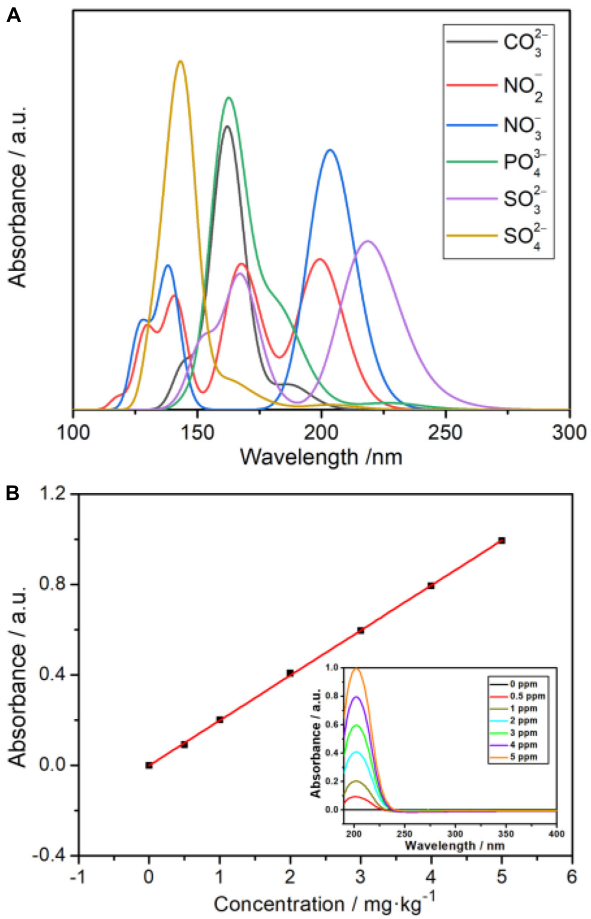


FIGURE 3 | (A) Absorbance of several ions in water from UV-Vis spectrophotometry. **(B)** Standard curves were made by UV-Vis spectrophotometry.

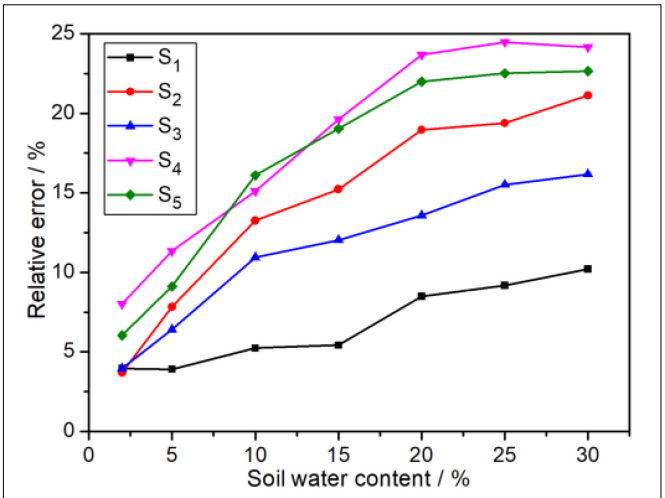


FIGURE 4 | Effects of soil water content on detection accuracy.

ignored. It was observed that the relative error (RE) of NO_3^--N grew up to 25% rapidly when the water content increases. Even if the soil water content is less than 2%, the RE is found to be greater than 3%. In addition, the RE toward soil water content is a slight difference, which may be caused by their distinguishing intrinsic properties.

As a result, it is necessary to provide a calibration function to remove the impact of inaccuracy caused by soil water content on concentration measurement uncertainty. The relationship is expressed as follows:

$$C_W = \left[\frac{C}{m} \times \left(1 + \frac{\omega}{100}\right) + \frac{\omega}{\rho_\omega \times 100}\right] \times \beta \qquad (2)$$

Where C_W denotes the calibrated NO_3^--N concentration (mg/kg), C denotes the volume of extraction solution (ml), m denotes the weight of the fresh soil sample (g), ω represents the soil water content (%),ρ_ω represents the density of water at the room temperature (1.00 g/cm^3), and β denotes the mass fraction of soil NO_3^--N that is calculated from the measured curve (mg/kg).

Table 1 displays the results from 10 soil samples with water contents ranging from 0 to 30%. The results of the ISE system were compared with the UV-Vis spectrophotometry. The RE was fluctuated up to −24.02% before the measurement findings were calibrated by the calibration function. However, after the water content function was added, it dropped to −10.93%. The RE was effectively reduced to 13.09%.

Performance Analysis

To evaluate this cooperative all-solid-state ISE system, the response time, which is defined as the time interval of electrode potential reaching 95% of the initial potential, was investigated. **Figure 5A** illustrates the change of the response time at the various logarithm of (NO_3^-). It can be observed that the steady potentials can be obtained within less than 15 s, which is comparable with the previous report. Notably, the response time shortens with increasing the (NO_3^-), implying that the higher (NO_3^-) accelerates the diffusion from the external solution to the internal solution driven by concentration difference and makes the electrode potential steady more quickly. In addition, we studied the linear response interval of electrode potential toward the logarithm of (NO_3^-). As shown in **Figure 5B**, the linear range of (NO_3^-) can be directly determined as 10^{-5}–$10^{-2.2}$ mol/L. Subsequently, we further studied the rationality of the measured electrode potential in the linear range *via* linear fitting by the Nernst equation of Eq. 2, as shown in **Figure 5C**. The slope is −51.63 mV/decade which is laid in the rational range of (−54 ± 5) mV/decade. The vertical intercept is 156.68 mV, which is related to E_0, pH, and T. Notably, recovery rates were also analyzed on all standard samples used in the above experiments, and the results were in an acceptable range of 90–110% (**Table 2**).

We explored the reliability of the cooperative all-solid-state ISE system toward UV-Vis spectrophotometer with an R^2 of 0.9952, as shown in **Figure 5D**. The fitted slope is 1.04, indicating the measured results from our ISE system almost reach to that of the UV-Vis spectrophotometer. Nevertheless, the fitted vertical intercept is −15.96 mg/kg, suggesting the existence of hysteresis

TABLE 1 | NO_3^--N concentrations in soil with different soil water contents.

Sample No.	Water content/%	NO_3^--N ISE system/(mg/kg)		NO_3^--N UV-Vis spectrophotometry/(mg/kg)	RE/%	
		Uncalibrated	Calibrated		Uncalibrated	Calibrated
1	0	45.37	45.37	49.35	−8.07	−8.07
2	3.72	50.59	52.85	52.24	−3.16	1.17
3	6.82	29.24	31.63	33.83	−13.59	−6.51
4	8.26	16.45	18.08	19.30	−14.77	−6.32
5	12.88	12.40	14.32	13.55	−8.50	5.64
6	16.62	11.05	13.26	12.82	−13.78	3.42
7	18.67	27.28	33.39	30.86	−11.60	8.20
8	22.91	6.55	7.68	8.62	−24.02	−10.93
9	27.38	30.65	40.72	37.11	−17.39	9.75
10	30.32	40.35	55.03	49.94	−19.21	10.19

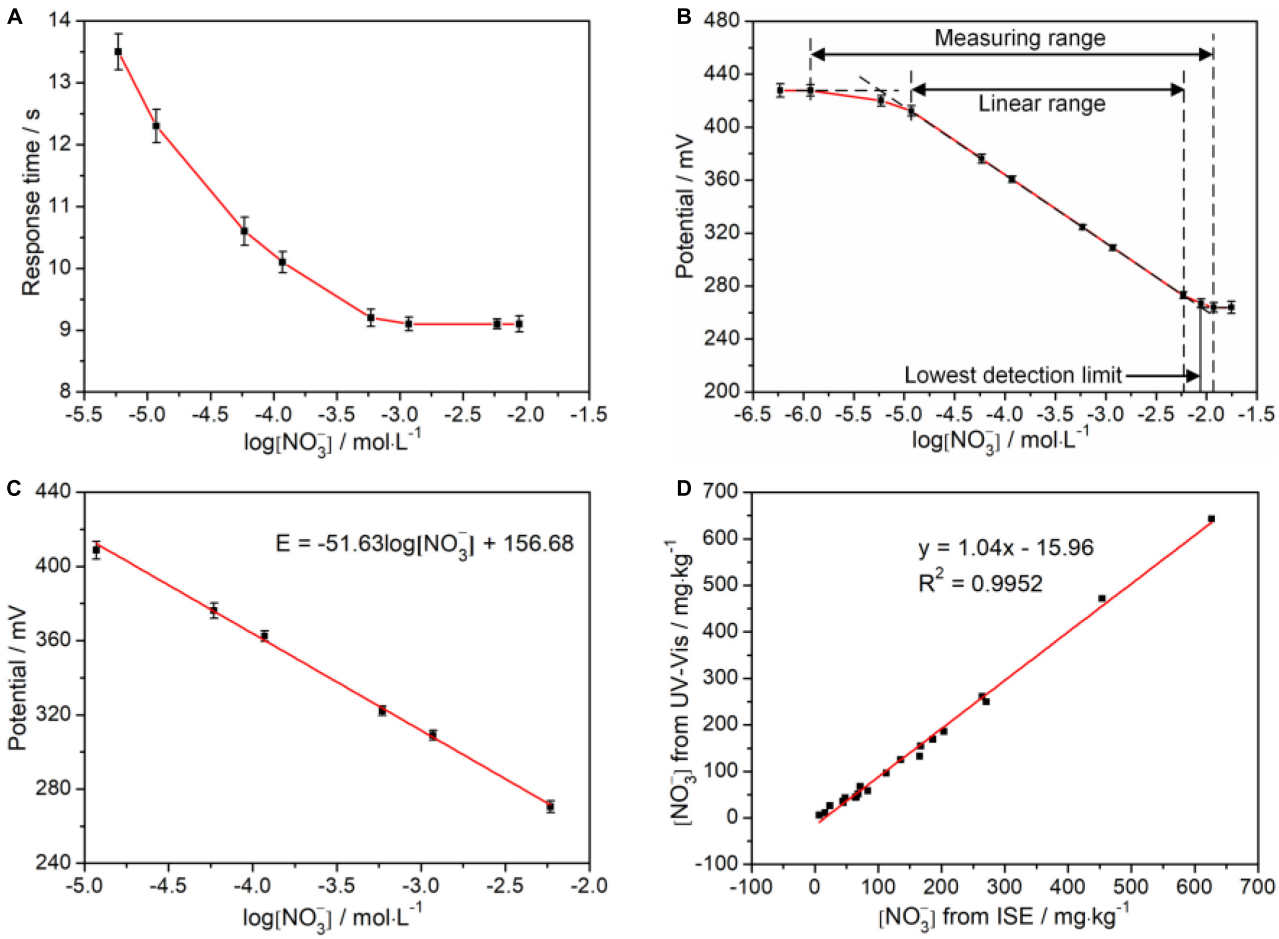


FIGURE 5 | Validation analyses of the ion-selective electrode (ISE) system. **(A)** Response time; **(B)** Linear range; **(C)** Response potential; and **(D)** Correlation between the results from ISE system and UV-Vis spectrophotometry. The (NO_3^-) used in **(A–C)** is derived from standard solution. The (NO_3^--N) used in **(D)** is derived from the soil sample solution.

effect resulting from the residual (NO_3^-) in each measurement. Each measurement was repeated six times with both methods, respectively. The relative SDs (RSDs) were calculated to verify the precision. From **Table 3**, the RSD of the cooperative ISE system is within 3%, higher than 1% of the UV-Vis spectrophotometry. In comparison with the non-contacting UV-Vis method, the contacting measurement of the ISE system inevitably causes liquid junction, wetting, and clean problem, which leads to

TABLE 2 | Recovery rate of the ion-selective electrode (ISE) system on detecting NO_3^--N.

Initial sample concentration[α] mg/kg	Added standard solution concentration[γ] mg/kg	Total concentration after adding standard solution mg/kg	Recovery rate/(%)
5	3	3.90	93.33%
5	10	7.96	109.20%
25	15	20.59	107.87%
25	50	36.72	96.88%
50	100	76.95	103.90%
100	200	147.64	97.64%
200	400	305.81	102.91%

[α] *The sample volume is 20 ml.*
[γ] *The added standard solution volume is 20 ml.*

TABLE 3 | Contrast analysis between the cooperative ISE system and UV-Vis spectrophotometry.

Sample No.	Cooperative ISE system		UV-Vis spectrophotometry	
	NO_3^--N mg/kg	RSD (%)	NO_3^--N mg/kg	RSD (%)
1	166.97	2.06	154.32	1.02
2	186.53	1.94	169.16	0.77
3	165.56	2.08	132.72	1.03
4	264.25	1.77	261.20	0.49
5	112.80	1.60	96.50	0.71
6	15.88	3.79	11.19	1.40
7	23.64	3.93	26.53	0.78
8	6.83	3.96	5.84	0.15
9	48.15	2.87	43.10	0.77
10	83.68	2.03	58.39	1.05
11	44.23	2.71	35.76	1.54
12	68.48	2.47	51.78	0.58
13	71.59	2.39	67.77	1.52
14	65.52	2.82	44.52	0.77
15	64.10	2.20	44.10	0.61
16	45.19	2.85	33.13	0.91
17	135.45	1.46	124.86	0.34
18	203.61	2.17	185.53	0.66
19	270.89	1.81	249.71	0.61
20	626.64	0.78	643.20	0.32
21	453.72	1.19	471.55	0.59
Average		2.33		0.79

the fluctuation of measured potentials and increases the RSD. Thus, such measurement repeatability of our ISE system is acceptable for use in the field. To further verify whether there is a significant difference in the precision of both methods, the *F*-test was conducted in ANOVA. As shown in **Table 4**, there was no statistically significant difference ($P < 0.05$) between the all-solid-state ISE system and UV-Vis spectrophotometry. This proves that the electrode system can achieve comparable and relatively accurate results that meet the requirements of NO_3^--N measurement in the field.

Furthermore, compared with other ISE detection methods previously reported (**Table 5**), the detection of NO_3^--N using ISEs had a wider linear range and a relatively low limit of detection (LOD), the linear range was wider than that of UV-Vis spectrophotometry detection of NO_3^--N, and the detection speed was faster. The detection method is simple to operate and does not require the participation of professionals, which is providing

TABLE 4 | *F*-test of significance in the regression analysis of the experiment results.

Project	Statistic	Project	Statistic	Project	Statistic
Regression equation	$y = 1.039x - 15.956$	Coefficient *a* test	$a = 1.039$	Intercept *b* test	$b = -15.956$
SSR	509777.523 v = 1	*t*-value (a)	62.713	*t*-value (b)	−4.560
SSE	2462.746 v = 19	*P*-value (a)	0.000	*P*-value (b)	0.000
F-value	3932.916	S(a)	0.017	S(b)	3.500
P-value	0.000	95% CI	1.004∼1.074	95% CI	−23.281∼−8.632
Test result	$P < 0.05$, refuse H0, accept H1, indicating that there is a significant linear relationship between the two detection methods			$P > 0.05$, accept H0, indicating that there is no significant difference between intercept and 0 value	

TABLE 5 | Performance comparison on NO_3^--N measurement with different techniques.

Detection methods and principle	Measurement ranges	Sensitivity/response times	Limit of detection (LOD)	R^2	References
DSM[a]	–	−58.2 mV/decade	–	0.41−0.51	Adamchuk et al., 2005
ISE (PVC+Hitachi)	0.11−109.8 mg/L	–	–	0.89	Kim et al., 2007
SNMS[b]	–	6 s	–	0.93	Sibley et al., 2008
ISE (Horiba B-342)	6.8−68 mg/L	–	2 mg/L	0.96	Tully and Weil, 2014
Cooperative ISEs	10^{-5}−$10^{-2.2}$ mol/L	<15 s	$10^{-5.23}$ mol/L	0.99	This work

[a] *A method of direct soil measurement (DSM) using ISEs.*
[b] *Soil nitrate mapping system.*

rational suggestions for relieving the leaching of NO_3^--N during fertilization and irrigation.

CONCLUSION

To detect NO_3^--N in the field, an electrode system consisting of all-solid-state NO_3^- ISEs, temperature electrode, and pH electrode was built. A microprocessor-controlled peristaltic pump extracted the measured sample solution automatically in the sample cell. The NO_3^--N concentration measured by the ISE system was quantitatively calibrated by adding soil water content calibration formula. The RE is effectively reduced to 13.09%. Four different fertilization treatments were carried out in experimental fields of crop growing areas in Henan and Hubei provinces, each area used a plum-shaped cloth point approach, and the cloth point was measured using the ISE system. In the measurement range of 10^{-5}–$10^{-2.2}$ mol/L, the ISE system has a response time of less than 15 s with a slope laid in the rational range of (-54 ± 5) mV/decade. The recovery rate of 90–110% has been confirmed from the ISE system for soil NO_3^--N. Both the RSDs of 3% of the soil NO_3^--N were obtained from this ISE system. Compared to the classical UV-Vis spectrophotometer, an R^2 of 0.9952 has been obtained. The linear regression F-test has been carried out, and there was a significant linear relationship between the measurement results of the two detection systems. We believe that this system can be further optimized and generalized in agriculture, providing rational suggestions for relieving the leaching of NO_3^--N during fertilization and irrigation.

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

RS: conceptualization, writing-original draft, and writing-review and editing. JW: validation and writing-review and editing. JH: conceptualization, validation, and writing-review and editing. LM: formal analysis and investigation. SA: writing-review and editing. YZ: formal analysis and supervision. MA and ZB: supervision. LL and WW: software and formal analysis. CL: data curation and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (No. 32071890) and supported by the Henan Center for Outstanding Overseas Scientists (No. GZS2021007).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.810214/full#supplementary-material

REFERENCES

Adamchuk, V. I., Lund, E. D., Sethuramasamyraja, B., Morgan, M. T., Dobermann, A., and Marx, D. B. (2005). Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. *Comput. Electron. Agric.* 48, 272–294. doi: 10.1016/j.compag.2005.05.001

Burton, L., Dave, N., Fernandez, R. E., Jayachandran, K., and Bhansali, S. (2018). Smart gardening IoT soil sheets for real-time nutrient analysis. *J. Electrochem. Soc.* 165, B3157–B3162. doi: 10.1149/2.0201808jes

Burton, L., Jayachandran, K., and Bhansali, S. (2020). Review—the "Real-Time" revolution for in situ soil nutrient sensing. *J. Electrochem. Soc.* 167:037569. doi: 10.1149/1945-7111/ab6f5d

Cho, W.-J., Kim, H.-J., Jung, D.-H., Kim, D.-W., Ahn, T. I., and Son, J.-E. (2018). On-site ion monitoring system for precision hydroponic nutrient management. *Comput. Electron. Agric.* 146, 51–58. doi: 10.1016/j.compag.2018.01.019

Cui, M., Zeng, L., Qin, W., and Feng, J. (2020). Measures for reducing nitrate leaching in orchards:a review. *Environ. Pollut.* 263:114553. doi: 10.1016/j.envpol.2020.114553

Gebbers, R., and Adamchuk, V. I. (2010). Precision agriculture and food security. *Science* 327, 828–831. doi: 10.1126/science.1183899

Kim, H.-J., Hummel, J. W., Sudduth, K. A., and Motavalli, P. P. (2007). Simultaneous analysis of soil macronutrients using ion-selective electrodes. *Soil Sci. Soc. Am. J.* 71, 1867–1877. doi: 10.2136/sssaj2007.0002

Li, X., Govind, N., Isborn, C., DePrince, A. E. III, and Lopata, K. (2020). Real-time time-dependent electronic structure theory. *Chem. Rev.* 120, 9951–9993. doi: 10.1021/acs.chemrev.0c00223

Ma, L., Li, Z., Birech, Z., Li, S., Yang, Y., Zhang, W., et al. (2019b). Multi-channel optoelectronic measurement system for soil nutrients analysis. *Electronics* 8:451. doi: 10.3390/electronics8040451

Ma, L., Duan, T., and Hu, J. (2019a). Application of a universal soil extractant for determining the available NPK: a case study of crop planting zones in central China. *Sci. Total Environ.* 704:135253. doi: 10.1016/j.scitotenv.2019.135253

Mahmud, M., Ejeian, F., Azadi, S., Myers, M., and Asadnia, M. (2020). Recent progress in sensing nitrate, nitrite, phosphate, and ammonium in aquatic environment. *Chemosphere* 259:127492. doi: 10.1016/j.chemosphere.2020.127492

Marenich, A. V., Cramer, C. J., and Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B.* 113, 6378–6396. doi: 10.1021/jp810292n

Mary, W., Rena, J., Jean, B., Theo, D. K., Peter, W., Bernard, N., et al. (2018). Drinking water nitrate and human health: an updated review. *Int. J. Environ. Res. Public Health* 15:1557. doi: 10.3390/ijerph15071557

Neese, F. (2011). The ORCA program system. *WIRES Comput. Mol. Sci.* 2, 73–78. doi: 10.1002/wcms.81

Nyameasem, J. K., Reinsch, T., Taube, F., Domozoro, C. Y. F., Marfo-Ahenkora, E., Emadodin, I., et al. (2020). Nitrogen availability determines the long-term impact of land use change on soil carbon stocks in grasslands of southern Ghana. *SOIL* 6, 523–539. doi: 10.5194/soil-6-523-2020

Pennino, M. J., Compton, J. E., and Leibowitz, S. G. (2017). Trends in drinking water nitrate violations across the United States. *Environ. Sci. Technol.* 51, 13450–13460. doi: 10.1021/acs.est.7b04269

Richa, A., Fizir, M., and Touil, S. (2021). Advanced monitoring of hydroponic solutions using ion-selective electrodes and the internet of things: a review. *Environ. Chem. Lett.* 19, 3445–3463. doi: 10.1007/s10311-021-01233-8

Rogovska, N., Laird, D. A., Chiou, C.-P., and Bond, L. J. (2018). Development of field mobile soil nitrate sensor technology to facilitate precision fertilizer management. *Precis. Agric.* 20, 40–55. doi: 10.1007/s11119-018-9579-0

Sadler, R., Maetam, B., Edokpolo, B., Connell, D., Yu, J., Stewart, D., et al. (2016). Health risk assessment for exposure to nitrate in drinking water from village wells in Semarang, Indonesia. *Environ. Pollut.* 216, 738–745. doi: 10.1016/j.envpol.2016.06.041

Sibley, K. J., Astatkie, T., Brewster, G., Struik, P. C., Adsett, J. F., and Pruski, K. (2008). Field-scale validation of an automated soil nitrate extraction and measurement system. *Precis. Agric.* 10, 162–174. doi: 10.1007/s11119-008-9081-1

Sun, J., Li, W., Li, C., Chang, W., and Peng, M. (2020). Effect of different rates of nitrogen fertilization on crop yield, soil properties and leaf physiological attributes in banana under subtropical regions of China. *Front. Plant Sci.* 11:613760. doi: 10.3389/fpls.2020.613760

Szpak, P. (2014). Complexities of nitrogen isotope biogeochemistry in plant-soil systems: implications for the study of ancient agricultural and animal management practices. *Front. Plant Sci.* 5:288. doi: 10.3389/fpls.2014.00288

Tully, K. L., and Weil, R. (2014). Ion-selective electrode offers accurate, inexpensive method for analyzing soil solution nitrate in remote regions. *Commun. Soil Sci. Plant Anal.* 45, 1974–1980. doi: 10.1080/00103624.2014.912297

Van Groenigen, J. W., Huygens, D., Boeckx, P., Kuyper, T. W., Lubbers, I. M., Rütting, T., et al. (2015). The soil N cycle: new insights and key challenges. *SOIL* 1, 235–256. doi: 10.5194/soil-1-235-2015

Velusamy, K., Periyasamy, S., Kumar, P. S., Vo, D., Sindhu, J., Sneka, D., et al. (2021). Advanced techniques to remove phosphates and nitrates from waters: a review. *Environ. Chem. Lett.* 19, 3165–3180. doi: 10.1007/s10311-021-01239-2

Wang, J., Lü, G., Guo, X., Wang, Y., Ding, S., and Wang, D. (2015). Conservation tillage and optimized fertilization reduce winter runoff losses of nitrogen and phosphorus from farmland in the Chaohu Lake region, China. *Nutr. Cycl. Agroecosyst.* 101, 93–106. doi: 10.1007/s10705-014-9664-3

Zhang, H., and Yao, Y. (2017). Vermiculite addition to soil decreases N water pollution by over 30%. *Environ. Chem. Lett.* 15, 507–513. doi: 10.1007/s10311-017-0631-5

Zhao, Y., and Truhlar, D. G. (2007). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* 120, 215–241. doi: 10.1007/s00214-007-0310-x

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fpls.2021.791256

Soybean Yield Preharvest Prediction Based on Bean Pods and Leaves Image Recognition Using Deep Learning Neural Network Combined With GRNN

Wei Lu[1], *Rongting Du*[1], *Pengshuai Niu*[1], *Guangnan Xing*[2], *Hui Luo*[1], *Yiming Deng*[3] *and Lei Shu*[1]*

[1] College of Artificial Intelligence, Nanjing Agricultural University, Nanjing, China, [2] College of Agriculture, Nanjing Agricultural University, Nanjing, China, [3] College of Engineering, Michigan State University, East Lansing, MI, United States

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Zhao Zhang,
North Dakota State University, United States
Bo Liu,
California Polytechnic State University, United States
Dandan Wang,
Xi'an University of Science and Technology, China

***Correspondence:**
Lei Shu
lei.shu@njau.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *08 October 2021*
Accepted: *08 December 2021*
Published: *13 January 2022*

Citation:
Lu W, Du R, Niu P, Xing G, Luo H, Deng Y and Shu L (2022) Soybean Yield Preharvest Prediction Based on Bean Pods and Leaves Image Recognition Using Deep Learning Neural Network Combined With GRNN. Front. Plant Sci. 12:791256. doi: 10.3389/fpls.2021.791256

Soybean yield is a highly complex trait determined by multiple factors such as genotype, environment, and their interactions. The earlier the prediction during the growing season the better. Accurate soybean yield prediction is important for germplasm innovation and planting environment factor improvement. But until now, soybean yield has been determined by weight measurement manually after soybean plant harvest which is time-consuming, has high cost and low precision. This paper proposed a soybean yield in-field prediction method based on bean pods and leaves image recognition using a deep learning algorithm combined with a generalized regression neural network (GRNN). A faster region-convolutional neural network (Faster R-CNN), feature pyramid network (FPN), single shot multibox detector (SSD), and You Only Look Once (YOLOv3) were employed for bean pods recognition in which recognition precision and speed were 86.2, 89.8, 80.1, 87.4%, and 13 frames per second (FPS), 7 FPS, 24 FPS, and 39 FPS, respectively. Therefore, YOLOv3 was selected considering both recognition precision and speed. For enhancing detection performance, YOLOv3 was improved by changing IoU loss function, using the anchor frame clustering algorithm, and utilizing the partial neural network structure with which recognition precision increased to 90.3%. In order to improve soybean yield prediction precision, leaves were identified and counted, moreover, pods were further classified as single, double, treble, four, and five seeds types by improved YOLOv3 because each type seed weight varies. In addition, soybean seed number prediction models of each soybean planter were built using PLSR, BP, and GRNN with the input of different type pod numbers and leaf numbers with which prediction results were 96.24, 96.97, and 97.5%, respectively. Finally, the soybean yield of each planter was obtained by accumulating the weight of all soybean pod types and the average accuracy was up to 97.43%. The results show that it is feasible to

predict the soybean yield of plants *in situ* with high precision by fusing the number of leaves and different type soybean pods recognized by a deep neural network combined with GRNN which can speed up germplasm innovation and planting environmental factor optimization.

Keywords: yield prediction, phenotyping, germplasm innovation, soybean, *in situ*

INTRODUCTION

Soybean is an important source of high quality protein and oil in the world, which contains about 42% protein, 20% oil, and 33% carbohydrate (Zhang et al., 2001). Soybean protein can enhance the body's immune function and play an important nutritional role in human tissues and cells. Soybean production for 2020 totaled 4.14 billion bushels, up 16% from 2019 in America due to a higher average soybean yield (Alsajri et al., 2020). Soybean yield prediction is of great importance to global food production, which is a highly complex trait determined by multiple factors such as genotype, environment, and their interactions. Accurate soybean yield prediction is important for germplasm innovation and planting environment factor improvement. Many researchers have tried to clarify the phenotype (such as yield) as an explicit function of the genotype (G), environment (E), and their interactions (G×E). In fact, the selection of individuals with good genotypic effect can further improve the yield of existing soybean varieties, which is also of great significance for high-yield soybean breeding. So, the earlier the prediction during the growing season the better. But until now, soybean yield has been popularly determined by manual weight measurement after soybean plant harvest which is time-consuming, expensive, and imprecise.

In recent years, digital image processing combined with machine learning technology has been applied for crop yield prediction in literature. The relationship between grain area and weight was studied using an image processing method (Zhao et al., 2019). A citrus fruit crop prediction algorithm based on color difference of citrus fruit and leaves was studied in citrus trees (Dorj et al., 2017). A region growing algorithm was proposed to segment cotton bolls into color images and count them and predict yield (Sun et al., 2019). Algorithms that rely on feature extraction (SVM, NN, RF) and algorithms that do not need feature extraction (GoogLeNet, VGG-16) were compared. The study found that the VGG-16 algorithm could effectively distinguish corn and soybean (Flores et al., 2020). In addition, a multi-rotor UAV system was developed to obtain high-resolution images and information related to geographical location, shooting angle, and environmental illumination, so as to carry out effective agricultural detection (Zhu et al., 2019).

More recently, deep neural networks have been employed in crop yield prediction, including the convolutional neural network (CNN), faster region-convolutional neural network (Faster R-CNN), single shot multibox detector (SSD), and You Only Look Once (YOLO), etc. The features of hyperspectral and color images was used to classify corn and estimate corn yield by CNN (Yang et al., 2021). Faster R-CNN has been modified for detection and yield estimation of fruits (mangoes, pomegranates, tomatoes, apples, and oranges) (Behera et al., 2021), which has also been used in prediction of melon yield (Zhao et al., 2017). Comparation of Faster R-CNN and SSD in citrus counting and yield prediction has been carried out (Qin et al., 2021). A convolutional neural network combined with linear regression was used in sorghum spike identification and weight prediction (Zannou and Houndji, 2019). Lightweight YOLO was applied to predict the yield of oil palm fruit based on images collected by UAV (Junos et al., 2021). The traditional rectangular bounding box in YOLOv3 was replaced as a circular rectangular box for better positioning of tomatoes (Liu et al., 2020). A deep neural network was designed to study the influence of genotype, environment, and their interaction on yield prediction (Khaki and Wang, 2019). There are also studies using satellite images to predict small yield using machine learning methods such as the Gaussian process regression algorithm (Sharifi, 2021).

Existing deep learning detection targets such as strawberries (Yu et al., 2019), tomatoes (Hu et al., 2019), apple (Tian et al., 2019), pepper (Hespeler et al., 2021), etc. are obviously different from the background leaves and branches of plants, which brings convenience for fruit identification due to remarkable color difference. Although the color of cucumbers (Mao et al., 2020) and corn (Jin et al., 2018) ears are similar to the leaves and vines, they have large size, small number, and low density, which also reduces the difficulty of identification. In the yield prediction of wheat (Yang et al., 2019) and rice (Crisóstomo de Castro Filho et al., 2020), the dense clustering of wheat and rice grains increases the difficulty of detection, but fortunately, the leaves shield the ears of wheat and rice slightly. A mature soybean phenotype measurement algorithm called soybean phenotype measure-instance segmentation was proposed to calculate pod length, pod width, stem length, seed length, and seed width based on PCA combined with CNN (Li et al., 2021). For soybean yield detection, leaves and pods have similar color, pods are blocked by numerous leaves, moreover the pods are clustered together, which creates a huge challenge for soybean yield forecast. Moreover, the types of pods must also be identified at the same time to predict soybean production accurately because the number of grains in different types of pods are different.

Deep neural networks, unlike early shallow neural networks with a single hidden layer, have multiple hidden layers which can effectively reveal the underlying unknown and highly non-linear relationship between the input data and output variables (LeCun et al., 2015), which have been widely employed in face recognition, automatic driving, etc., but they also require more hardware and time consumption. Generally, more neural

network layers and nodes means the network is more powerful and paradoxically needs more time and hardware to run.

After comparing several mainstream deep neural networks, the YOLOv3 algorithm was chosen to recognize soybean pods and leaves in this paper. Moreover, in order to further improve the detection performance of the neural network, structure improvements were made to YOLOv3 to achieve accurate recognition of large leaves and small pods simultaneously. In addition, a generalized regression neural network (GRNN) model was established for prediction of seed number in a soybean plant by using the cumulative results of leaves and different type pods of four images taken at 90-degree intervals from different directions. Finally, soybean plant production was calculated based on the average grain weights of different type pods, which provided a new method and solution for soybean phenotype detection and germplasm innovation acceleration.

The remainder of this paper is organized as follows. Section Materials and Methods describes the materials and data used in this research and offers a detailed description of our improved deep neural network for soybean pods and leaves prediction, and yield modeling as well. Section Results and Discussion presents the results of our algorithm and models. Finally, the conclusion is given in section Conclusion.

MATERIALS AND METHODS

In this paper, the soybean yield prediction method included three steps as shown in **Figure 1**. First, the original imaging data of soybean plants collected were preprocessed by filtering and enhancement. Then the improved YOLOv3 model was used to identify and count leaves and different type soybean pods including one bean pods, two bean pods, three bean pods, four bean pods, and five bean pods. Finally, a GRNN model for predicting soybean yield was established based on the numbers of leaves and all categories of pods as inputs.

Materials and Image Sampling

In this study, soybean plants were grown in pots in Wanjiang Experimental Station of Nanjing Agricultural University where the latitude and longitude are 118.62°E and 31.54°N, respectively. Each pot had four soybean plants which were evenly distributed in the pot. There were 90 pots, 360 plants, and 24 varieties in total (Xudou-18, Heidou-2, Erzaohuang, Qiyuehuang-1, Bayuehuang-4, Qingyuanxiaoqingdou, Fengdudahuangdou, Enwangheidou, Kaijiangdongdou, Huazhouhuangdou, Ganyulianmaoshao, Liyangmaojiajia, Andingxiaoheidou, Nannong1606, P06, P12, P23, P25, P53, P59, P65, GS171761, GS71244, GS71411) used for the study. All potted soybean plants were at the pod-setting stage during image sampling.

A camera (model: Intel RealSense D435 manufactured by Intel) was employed to capture soybean plant images at three different time periods (6:00–7:00, 13:00–14:00, 17:00–18:00) in order to realize the completeness of images in different light intensity environments. The camera was 1.2 m above the ground and 1.5 m horizontally from the target while image capturing. Each pot of soybean plants was photographed in four different directions at intervals of 90 degrees, so a total of 360 images were collected.

Data processing was performed on a computer with a Win10 operating system, Intel(R) Core(TM) i7-8750H processor, 8GB memory, and a NVIDIA GeForce GTX 1050 Ti display adapter. The open source deep learning framework Tensorflow2.0 was used to establish models.

Soybean Pods and Leaves Recognition

Image Processing

In order to extract clear soybean plant images, denoising and enhancement treatments were carried out on soybean plant pictures due to the light environment and haze which blurred the pictures. First, *guide filter* was selected for denoising images after comparing them with the *bilateral filter* (Yu et al., 2020) and *DWT* (Rai et al., 2012) algorithms. Then, the *gamma* algorithm was chosen for image enhancement by contrast with *laplus* (Bhairannawar et al., 2017) and *log* (Maini and Aggarwal, 2010) algorithms. The pseudocode is shown in **Figure 2**.

Since deep neural networks require a certain number of training sets to improve the accuracy of the model, a data augmentation method, a common technique in deep learning research, was employed to increase the number of existing

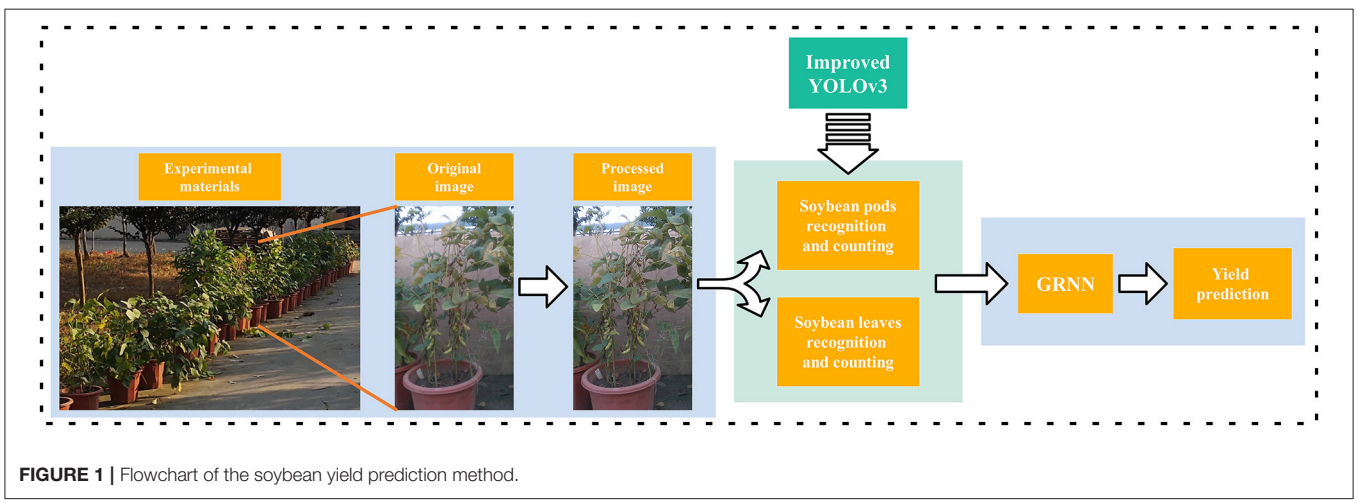


FIGURE 1 | Flowchart of the soybean yield prediction method.

Input: filtering input image p,guidance image I,radius r, regularization ϵ, control parameter c、γ.

Output: filtering output q. enhancement output s.

Guided Filter:

$$mean_I = f_{mean}(I)$$
$$mean_p = f_{mean}(p)$$
$$corr_I = f_{mean}(I.*I)$$
$$corr_{I_p} = f_{mean}(I.*p)$$
$$var_I = corr_I - mean_{I.} * mean_I$$
$$cov_{I_p} = corr_{I_p} - mean_{I.} * mean_p$$
$$a = cov_{I_p.}/(var_I + \epsilon)$$
$$b = mean_p - a.* mean_I$$
$$mean_a = f_{mean}(a)$$
$$mean_b = f_{mean}(b)$$
$$q = mean_{a.} * I + mean_b$$

Gamma Transformation:

$$s = c * q^{\gamma}$$

FIGURE 2 | The pseudocode of denoising and enhancement algorithms.

photos to 1,800 images by rotation, scaling, mirroring, random brightness increases/decreases, and other methods. The 1,800 images were randomly divided into the training set and test set according to the ratio of 4:1, which were then used as the dataset of target detection models. The image processing process is shown in **Figure 3**.

Soybean Leaves, Pods, and Types Recognition

With the rapid improvement of the computing power of computers, deep learning has made tremendous progress, and a lot of target detection algorithms based on deep learning have been proposed. Popular deep learning algorithms such as Faster R-CNN (Ren et al., 2016), feature pyramid network (FPN, Lin et al., 2017a), SSD (Liu et al., 2016), and YOLO (Redmon et al., 2016) have been applied in different areas and show very superior performance. So, these algorithms were utilized in soybean pod and type recognition in the paper.

The detection of Faster R-CNN includes two steps, the first step is region proposal network (RPN). Features of a picture are extracted by a VGG-16 (Simonyan and Zisserman, 2014) neural network, followed by foreground background classification and first prediction of the coordinate of the generated anchor. In the second step, candidate boxes with higher confidence are selected and sent to the back of the network for the second prediction

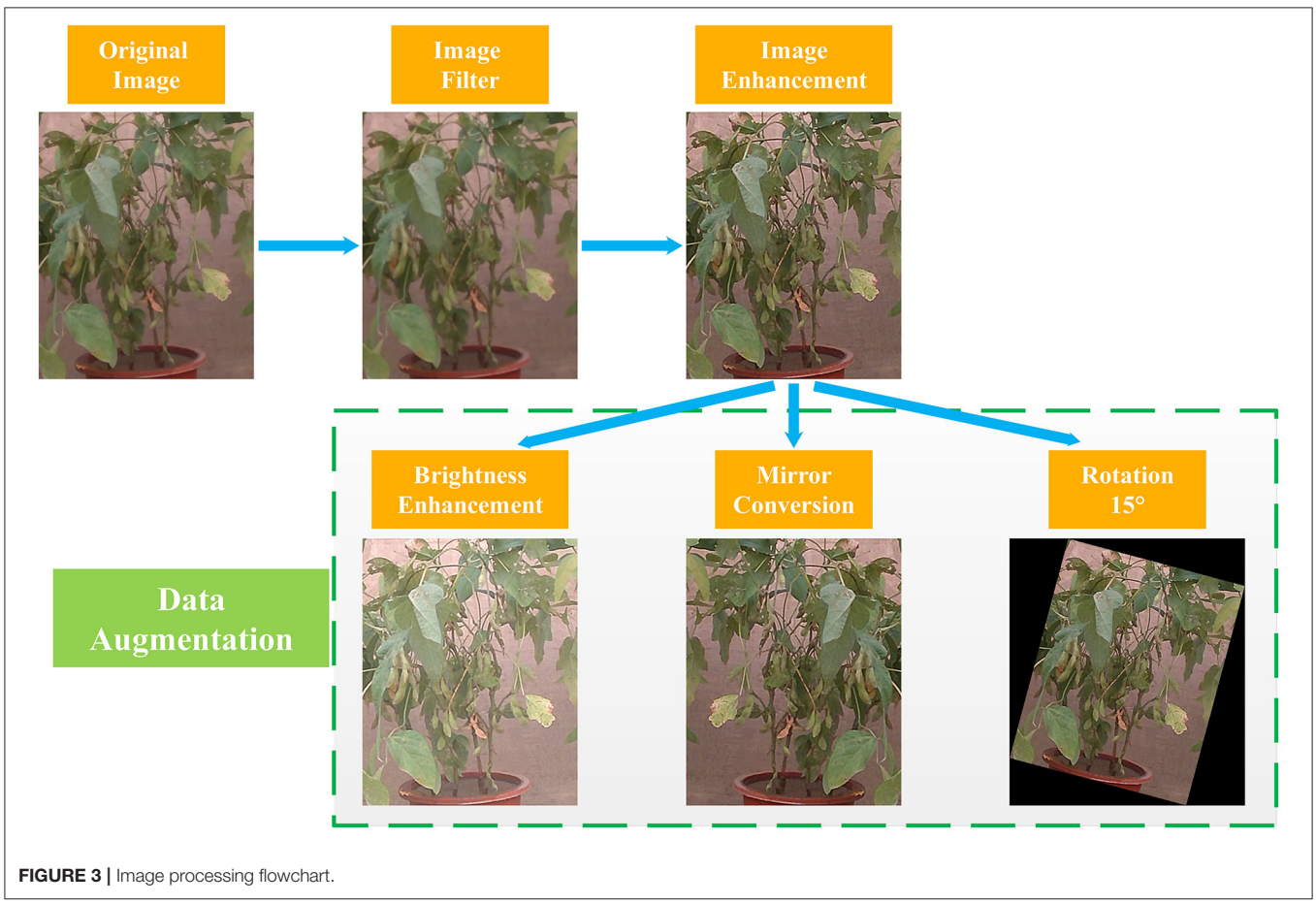


FIGURE 3 | Image processing flowchart.

of category and coordinate values, so as to predict the specific category. Faster R-CNN has extremely high prediction accuracy but is time-consuming for training (Benjdira et al., 2019).

The feature pyramid network (FPN) is proposed to alleviate the problems of multi-scale and small target detection which selects ResNet50 (He et al., 2016) as the feature network, and fuses the high-level features with the low-level features through the up-sampling process. Moreover, layers of the same class have horizontal connections and each layer is predicted independently. Therefore, the network has more abundant features. In addition, the idea of multi-scale detection is introduced into the RPN, and the anchor frame can be generated in different scales to cover different sized objects. FPN has good accuracy and precision but again, is more time-consuming.

In the single shot multibox detector (SSD) model, there is only one detection process step. The model uses VGG as the feature extraction network, and uses multiple feature layers to predict the target. As a result, the whole model is more lightweight with both good detection accuracy and speed.

The design concept of YOLO and SSD is similar, both are one stage target detection algorithms for reaching fast performance. In YOLOv3 (Redmon and Farhadi, 2018), a 28M DarkNet-53 is employed in which its parameters are only half that of ResNet101 (Lin et al., 2017b), but the performance is close to it. Pre-clustering of YOLOv2 (Redmon and Farhadi, 2017) is inherited in YOLOv3 on the anchor frame for targets clustering in the data using nine anchor frame scales that mostly fit the targets. Moreover, the feature processing is carried out in three different levels by introducing the multi-scale concept of FPN. And multiple binary classifiers are used in the calculation of classification loss to avoid competition within the class.

There are some problems such as error detection and missing detection by using YOLOv3 for soybean leaves and pods recognition because the color of the pods is similar to the leaves and the detection performance of YOLOv3 for small target like pods is not ideal. According to the above problems, an improved approach of YOLOv3 was proposed by changing the network structure and changing the clustering algorithm for increasing prediction without sacrificing too much speed. Firstly, the feature map after one down-sampling session is superimposed to the input of the second and third residual blocks so as to increase the detail information in the deep feature map for detection. The improved network of YOLOv3 is shown in **Figure 4**. DBL stands for Darknetconv2D_BN_Leaky, and resn stands for the number of Res_units contained in the Res_block.

In the target detection algorithm, the anchor frame can help the model to fit the coordinate points of the target, so that the positioning task can be transformed from finding the position of the target in the image to learning the coordinate offset of the anchor frame relative to the target. In consideration of the fact that the traditional K-means algorithm is easy to implement, but the initial clustering center needs to be set artificially, in addition, different centers have great influence on the clustering results, the K-means++ algorithm (Arthur and Vassilvitskii, 2006) was selected to cluster the length and width of the anchor frame to make it fit the soybean pod and leaves better. The clustering used Euclidean distance as a metric and set k from 6 to 11 step by step. Because a large value of k reduced the convergence rate of the model, k was set to 9 in this study after several attempts. The clustering centers obtained after the convergence of the model were: (23,31), (27,21), (41,52), (64,42), (78,90), (76,103), (91,94), (106,97), and (113,107).

In terms of loss function, traditional YOLOv3 uses IoU Loss, which is composed of coordinate regression loss, confidence loss, and classification loss, and its calculation formula is shown in Equation 1.

$$L_{IoU} = 1 - IoU \tag{1}$$

IoU Loss has the characteristics of scale invariance. Although IoU Loss is more advantageous than mean square error, when the relationship of the prediction box and the real box is contained and being contained, IoU Loss will be a fixed value, which has great influence on the detection effect. When the two do not intersect, the value of IoU Loss is 0, which cannot be optimized. According to the above problems, DIoU Loss (Distance IoU Loss) was used to replace IoU Loss in traditional YOLOv3, which

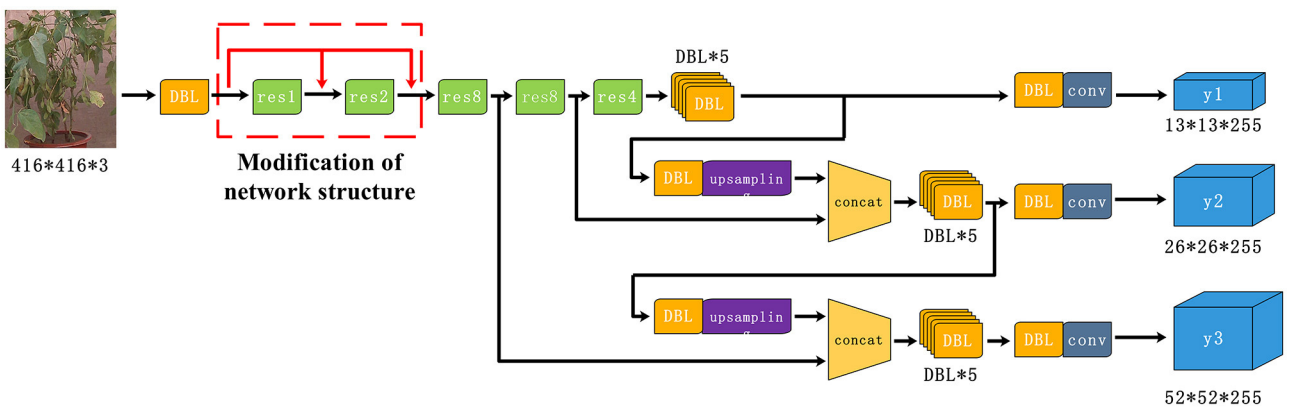


FIGURE 4 | Improved YOLOv3 for soybean leaves, pod, and type identification.

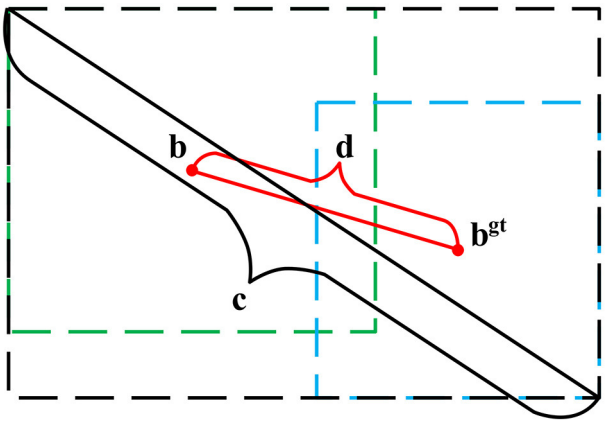


FIGURE 5 | Schematic diagram of DIoU.

made the regression of the target box more stable. The schematic diagram is shown in **Figure 5**.

DIoU is defined as,

$$DIoU = IoU - \frac{\rho^2\left(b, b^{gt}\right)}{c^2} \tag{2}$$

where b is the central coordinate of the prediction box, b^{gt} is the center coordinate of the real box, ρ is the Euclidean distance of the two center points, and c is the diagonal length of the minimum outer rectangle of the two target bounding boxes.

The final definition of the loss function DIoU Loss is shown in Equation 3.

$$L_{DIoU} = 1 - DIoU \tag{3}$$

In this paper, the improved YOLOv3 model was used to train the enhanced dataset. The size of network input was 416*416*3. The batch size was 64, the value of subbatch was 16, the momentum of dynamic parameter was 0.9, the maximum number of iterations was 14,000, the learning rate strategy was step decreasing, the initial value was 0.001, the scale parameter was 0.1, and the two step values of learning rate change were 11,200 and 12,600, respectively.

Soybean Yield Prediction

In the soybean yield prediction task based on plant images, the counting accuracy of pods was affected because some pods were occluded by leaves, and the density of leaves was positively correlated with the number of occluded pods. In addition, pods were divided into several types due to the difference number of seeds in them. The improved YOLOv3 model was used to identify the number of leaves and different types of pods, then PLSR (Geladi and Kowalski, 1986), BPNN (Hecht-Nielsen, 1992), and GRNN (Specht, 1991) models were established respectively to predict the amount of seeds. Among them, GRNN proposed by Specht has a strong non-linear mapping ability and learning speed which is an improved technique in neural networks based on non-parametric regression. It can even obtain good prediction accuracy but only requires a small number of datasets (Izonin et al., 2021). Moreover, the network can also handle unstable data in the inputs, which is suitable for soybean yield prediction. The architecture of GRNN is shown in **Figure 6**.

The mathematical representation is as below,

$$Y(x) = \frac{\sum_{k=1}^{N} y_k K(x, x_k)}{\sum_{k=1}^{N} K(x, x_k)} \tag{4}$$

where input x includes the numbers of soybean leaves and different type pods, $Y(x)$ is the predicted value of soybean yield, y_k is the activation weight for the pattern layer neuron at k, and $K(x, x_k)$ is the radial basis function kernel (Gaussian kernel) as formulated below.

$$K(x, x_k) = e^{-d_k/2\sigma^2} \tag{5}$$

$$d_k = (x - x_k)^T (x - x_k) \tag{6}$$

where d_k is the squared Euclidean distance between the training samples x_k and the input x.

Evaluation Indices

To evaluate the performance and stability of the proposed model, parameters such as prediction precision, recall, and degree of integration (IoU) were defined. Evaluation of the performance of the detection and recognition model is an essential stage. The detection accuracy and complexity are the key performance indexes in the evaluation. The basic evaluation indexes of the target detection model are accuracy rate (P) and recall rate (R). The definitions of the two indicators are shown in Equations 7, 8. Accuracy rate represents the ratio that the detected target really belongs to this class, which is used to describe the credibility of the target. Recall rate represents the ratio of the detected target to the actual total amount of the target, which is used to describe the degree of the target being found completely.

$$P = \frac{T_P}{T_P + F_P} \tag{7}$$

$$R = \frac{T_P}{T_P + F_N} \tag{8}$$

T_P–The number of correctly predicted targets
F_P–The number of wrongly predicted targets
F_N–The number of missed predicted targets

Mean average precision is used to measure the overall effect of multi-classification detection by averaging the detection precision of all categories. Its definition is shown in Formula 3.

$$MAP = \frac{\sum_{k=1}^{N} P(k)}{N} \tag{9}$$

Intersection over Union (IoU) is introduced to measure the similarity between the prediction box and real box according to the characteristics of the target detection task. Its definition is shown in Formula 4. When the prediction box is exactly

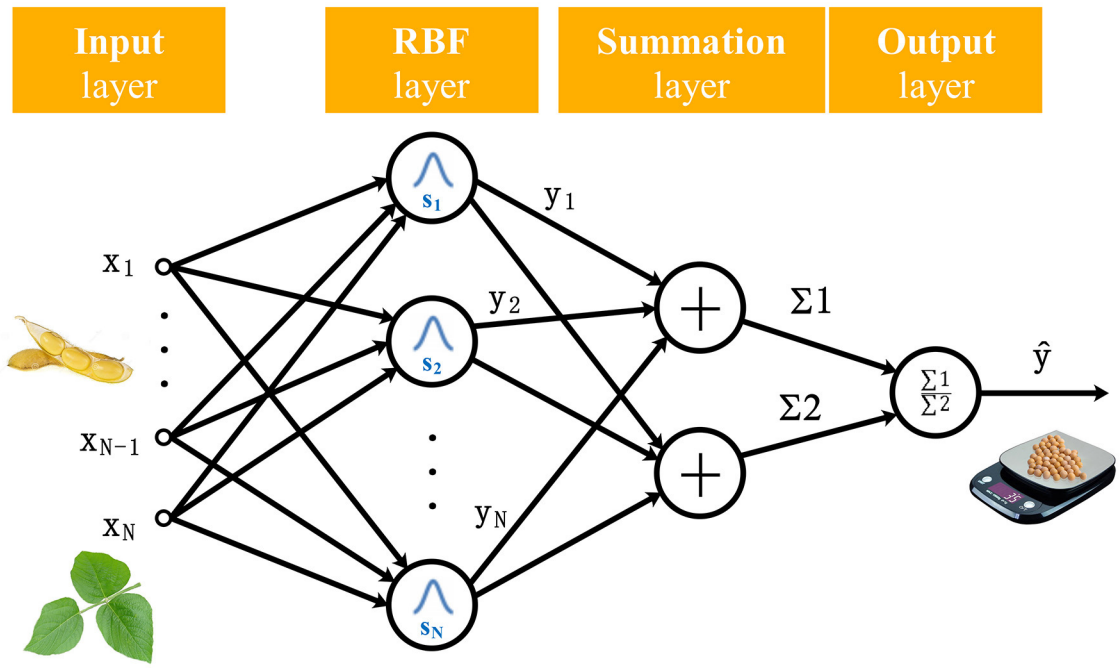


FIGURE 6 | Architecture of soybean yield prediction model of GRNN.

consistent with the real box, IoU is 1. Generally, the target is considered successfully detected when IoU is >0.5.

$$IoU = \frac{S_A \cap S_B}{S_A \cup S_B} \tag{10}$$

Soybean detection and recognition is a dichotomous problem, which only involves foreground soybean and background soybean. Therefore, F_1 was introduced to evaluate the model accuracy comprehensively. The value of F_1 depends on the accuracy and recall. Its definition is shown in Equation 11.

$$F_1 = \frac{2PR}{P+R} \tag{11}$$

The running speed of the detection and recognition algorithm is also an important basis for model evaluation. FPS was adopted as the evaluation standard in this experiment. Its definition is shown in Equation 12. In Equation 12, N represents the total number of samples and T represents the running time.

$$FPS = \frac{N}{T} \tag{12}$$

RESULTS AND DISCUSSION

Soybean Leaves and Pod Types Recognition Results

The performance of popular existing algorithms such as Faster R-CNN, FPN, SSD, and YOLOv3 in soybean pod detection tasks is shown in **Table 1**.

TABLE 1 | Soybean pod detection performance of different models.

Model		Accuracy *P*	Recall *R*	F_1 value	Speed
Existing algorithms	Faster R-CNN	86.2%	80.5%	83.3%	13FPS
	FPN	**89.8%**	82.7%	86.1%	7FPS
	SSD	80.1%	74.2%	77.0%	24FPS
	YOLOv3	87.4%	81.6%	84.4%	**39FPS**
Improved YOLOv3		**90.3%**	**87.6%**	**88.9%**	36FPS

The bold values are the best results.

The table indicates that FPN showed obvious advantages over the existing popular algorithms in prediction accuracy and *F*1 index up to 89.8% and 86.1% due to its relatively large structure, followed by YOLOv3, Faster R-CNN, and SSD with 87.4, 86.2, and 80.1% in accuracy and 84.4%, 83.3, and 77.0 in *F*1, respectively. YOLOv3 had the fastest speed at 39 FPS, followed by SSD, Faster R-CNN, and FPN with 24 FPS, 13 FPS, and 7 FPS, respectively. Among them, the speed of YOLOv3 was more than five times that of FPN, but the accuracy was 2.4% lower. Therefore, YOLOv3 was the best algorithm considering the prediction accuracy and speed comprehensively.

The accuracy of the improved YOLOv3 algorithm was 3.32% higher than that of YOLOv3, reaching 90.3%, and also better than that of FPN, SSD, and Faster R-CNN. The speed was second only to YOLOv3, but significantly faster than SSD, Faster R-CNN, and FPN, reaching 36 FPS, which can meet the demand of real-time recognition.

Moreover, considering that the number of seeds in different pods varied greatly, the pods were further classified into single

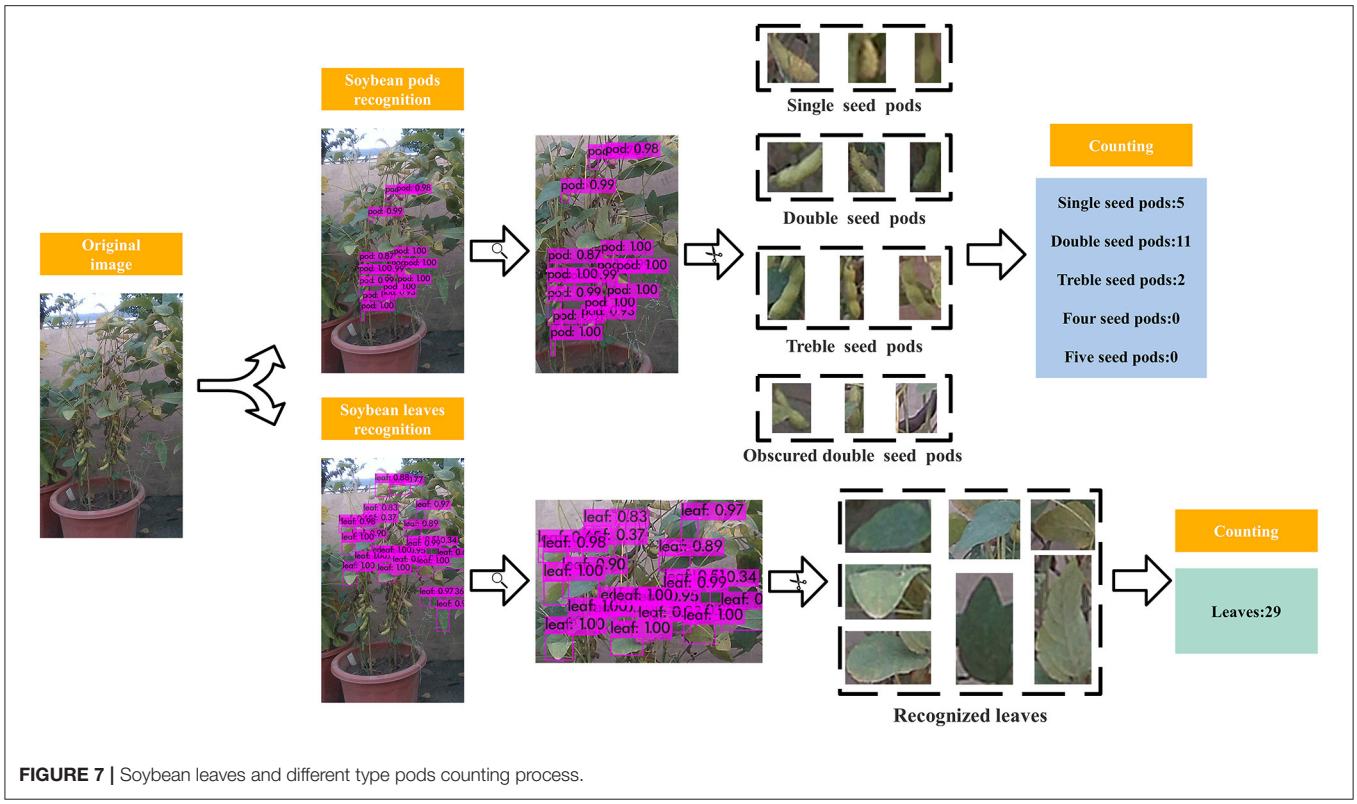


FIGURE 7 | Soybean leaves and different type pods counting process.

seed pods, double seeds pods, treble seeds pods, four seeds pods, and five seeds pods according to the number of internal seeds using the above proposed YOLOv3 model.

In order to accurately identify the pod type, pod type labeling and training were carried out on the recognized pod output anchor frame, and an accurate pod type recognition model was obtained. Soybean leaves and different type pods recognition and the counting process is shown in **Figure 7**.

TABLE 2 | Prediction accuracy of different models.

Model	*ACC*1	*ACC*2	*ACC*3	$\overline{ACC}$
PLSR	95.57%	96.24%	95.76%	95.84%
BPNN	96.57%	96.97%	96.59%	96.71%
GRNN	97.24%	97.50%	97.20%	**97.31%**

The bold values are the best results.

Soybean Plant Yield Prediction Results

In the prediction of soybean yield, taking the non-uniform characteristics of leaves growth into account, four pictures were taken from four directions of soybean plants at 90 degrees apart, and the total number of leaves and the total number of different types of pods of the four images were extracted by the above mentioned improved YOLOv3 algorithm, and then the seed number prediction models of PLSR, BPNN, and GRNN were established with the input of leaf number and different type pods number, of which the results shown in **Table 2** indicate that the GRNN model had the highest prediction accuracy. **Figure 8** shows the comparison between the actual yield of soybean per pot and the predicted yield when the GRNN model was used for yield prediction. After the model was run three times, the average accuracy of the GRNN model was up to 97.31%.

As shown in Equation 13, y_i and y'_i represent the actual value and predicted value of soybean seeds respectively, and n stands for the number of pots of all soybeans. *ACC* is the accuracy calculated according to the deviation degree between the predicted value and the actual value of soybean seed number, which can reflect the performance of the prediction models.

$$ACC = \frac{\sum_{i=1}^{n} \left(y_i - \left|y_i - y'_i\right|\right)/y_i}{n} \tag{13}$$

To further measure the weight of soybean grains produced by plants, the average weight of 100 soybean seeds ($w_a = 0.203$ g) and the average weight of 100 soybean seeds for each pod type 5 to 10 days after soybean harvesting from random sampling were measured. Among them, the average weight of soybean grains in single seed pods w_1, double grains pods w_2, treble grains pods w_3, four grains pods w_4, and five grains pods w_5 were 0.242 g, 0.207 g, 0.196 g, 0.189 g, and 0.186 g, respectively. Based on the numbers of different type pods predicted by the improved YOLOv3, the total weight of soybean grains produced was given, as shown in **Table 3**.

The accuracy of soybean yield predicted by GRNN was 95.14% with the average weight of 100 grains, while the accuracy of

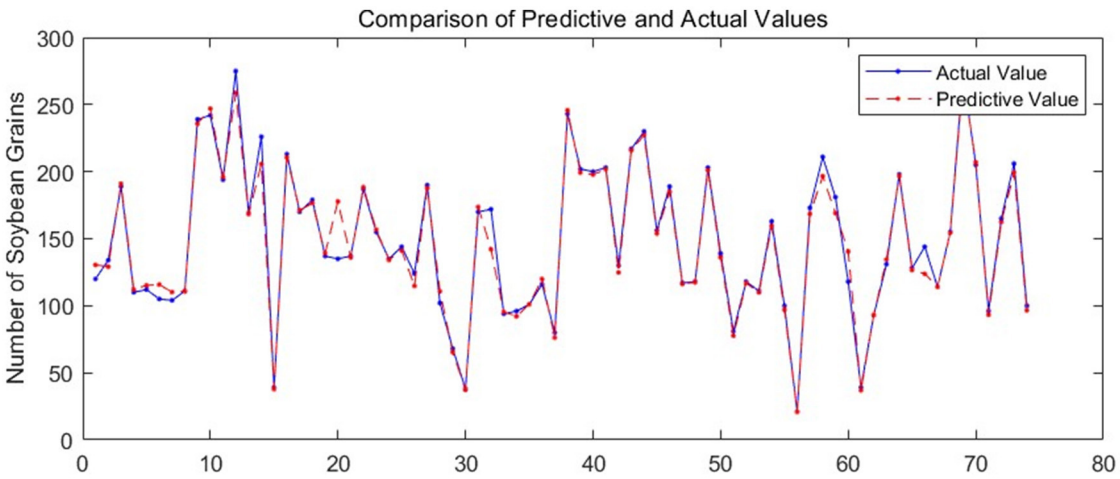


FIGURE 8 | Soybean seeds number prediction results of different plants.

TABLE 3 | Weight of single soybean grain of different pod types and total yield prediction results.

Pod type	Single (w_1)	Double (w_2)	Treble (w_3)	Four (w_4)	Five (w_5)	All types of pods (w_a)
Average weight per grain (g)	0.242	0.207	0.196	0.189	0.186	0.203
Total weight (g)			2668.670			2605.911
Accuracy			**97.43%**			95.14%

The bold values are the best results.

soybean yield predicted by GRNN with average weights of five different pods (w_1, w_2, w_3, w_4, and w_5) increased to 97.43%. The results show that it is more accurate to predict soybean yield by identifying and classifying soybean pods combined with GRNN.

CONCLUSION

Most of the existing crop yield prediction methods studied the impact of environmental changes on yield, but paid no attention to the actual yield prediction. Therefore, the most used field yield measurement method is still the traditional manual sampling process for statistical calculation, which is inefficient with low precision. Due to the high cost, time-consuming, and low accuracy of the traditional manual soybean yield measurement approach, this paper proposed a soybean yield *in situ* prediction method based on bean pods and leaves image recognition using a deep learning algorithm combined with a generalized regression neural network (GRNN). YOLOv3 is generally superior to Faster R-CNN, FPN, and SSD in terms of prediction accuracy and speed. Moreover, YOLOv3 was improved by changing the IoU loss function, using the anchor frame clustering algorithm, and utilizing the partial neural network structure in which recognition precision increased by 2.9% up to 90.3% at 36 FPS.

In this paper, we proposed to take four images of soybean plants at 90° intervals, and extract the total numbers of leaves and different type pods from the four images by improved YOLOv3. Then we established the prediction model of different type pods quantity of each plant using GRNN with inputs of the total numbers of leaves and different type pods recognized, in which the average accuracy increased to 97.31%, which was better than PLSR and BPNN. Furthermore, the soybean grain yield was calculated using the number and average weight of each type of pod. The prediction accuracy of the yield weight was up to 97.43%, which was better than the prediction accuracy based on the total number of grains and the average weight of different type pod grains.

This study shows that the improved YOLOv3 algorithm can be used to identify the number of leaves and different type pods and, moreover, can achieve accurate soybean yield *in situ* prediction 30–40 days in advance combined with the average weight of different soybean pods, which provides a new solution for accelerating soybean germplasm innovation and phenotypic detection of other crops.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

WL proposed the conceptualization and methodology and wrote the paper. RD programmed the software. PN compared the performance of the algorithms. GX designed and carried out the experiments. HL and LS improved the methodology and conceived the experiments. All authors reviewed the manuscript.

FUNDING

This research was funded by the National Natural Science Foundation of China (Nos. 32071896 and 31960487), the Agricultural Science and Technology Innovation Project of Jiangsu Province [No. CX(20)3068], the Modern Agricultural Machinery Equipment and Technology Demonstration and Promotion Project of Jiangsu Province (NJ2021-37), the Agricultural Science and Technology Innovation Project of Suzhou City (No. SNG2020039), and National Foreign Experts Program of China (G2021145010L).

REFERENCES

Alsajri, F. A., Wijewardana, C., Irby, J. T., Bellaloui, N., Krutz, L. J., and Golden, B. (2020). Developing functional relationships between temperature and soybean yield and seed quality. *Agron. J.* 112, 194–204. doi: 10.1002/agj2.20034

Arthur, D., and Vassilvitskii, S. (2006). *k-means++: The Advantages of Careful Seeding*. Stanford: Stanford University.

Behera, S. K., Rath, A. K., and Sethy, P. K. F. (2021). Yield estimation using Faster R-CNN with MIoU. *Multimed. Tools Appl.* 80, 19043–19056. doi: 10.1007/s11042-021-10704-7

Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., and Ouni, K. (2019). "Car detection using unmanned aerial vehicles: comparison between faster r-cnn and yolov3," in *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)* (Muscat: IEEE), 1–6.

Bhairannawar, S., Patil, A., Janmane, A., and Huilgol, M. (2017). "Color image enhancement using Laplacian filter and contrast limited adaptive histogram equalization," in *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)* (Vellore: IEEE), 1–5.

Crisóstomo de Castro Filho, H., Abílio de Carvalho Júnior, O., Ferreira de Carvalho, O. L., Pozzobon de Bem, P., Dos Santos de Moura, R., and Olino de Albuquerque, A. (2020). Rice crop detection using LSTM, Bi-LSTM, and machine learning models from sentinel-1 time series. *Remote Sens.* 12:2655. doi: 10.3390/rs12162655

Dorj, U. O., Lee, M., and Yun, S. (2017). An yield estimation in citrus orchards via fruit detection and counting using image processing. *Comput. Electron. Agric.* 140, 103–112. doi: 10.1016/j.compag.2017.05.019

Flores, P., Zhang, Z., Mathew, J., Jahan, N., and Stenger, J. (2020). Distinguishing volunteer corn from soybean at seedling stage using images and machine learning. *Smart Agric.* 2:61–74. doi: 10.12133/j.smartag.2020.2.3.202007-SA002

Geladi, P., and Kowalski, B. R. P. (1986). Least-squares regression: a tutorial. *Anal. Chim. Acta* 185, 1–17. doi: 10.1016/0003-2670(86)80028-9

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* (San Juan, PR), 770–778.

Hecht-Nielsen, R. (1992). *Theory of the Backpropagation Neural Network[M]//Neural Networks for Perception*. Cambridge, MA: Academic Press, 65–93.

Hespeler, S. C., Nemati, H., and Dehghan-Niri, E. (2021). Non-destructive thermal imaging for object detection via advanced deep learning for robotic inspection and harvesting of chili peppers. *Artif. Intell. Agric.* 5, 102–117. doi: 10.1016/j.aiia.2021.05.003

Hu, C., Liu, X., Pan, Z., and Automatic, L., i. P. (2019). detection of single ripe tomato on plant combining faster R-CNN and intuitionistic fuzzy set. *IEEE Access* 7, 154683–154696. doi: 10.1109/ACCESS.2019.2949343

Izonin, I., Tkachenko, R., Zub, K., and Tkachenko, P. (2021). A GRNN-based approach towards prediction from small datasets in medical application. *Procedia Comput. Sci.* 184, 242–249. doi: 10.1016/j.procs.2021.03.033

Jin, S., Su, Y., Gao, S., Wu, F., Hu, T., and Liu, J. (2018). Deep learning: individual maize segmentation from terrestrial lidar data using faster R-CNN and regional growth algorithms. *Front. Plant Sci.* 9:866. doi: 10.3389/fpls.2018.00866

Junos, M. H., Khairuddin, A. S. M., Thannirmalai, S., and Dahari, M. (2021). Automatic detection of oil palm fruits from UAV images using an improved YOLO model. *Vis. Comput.* 1–15. doi: 10.1007/s00371-021-02116-3

Khaki, S., and Wang, L. (2019). Crop yield prediction using deep neural networks. *Front. Plant Sci.* 10:621. doi: 10.3389/fpls.2019.00621

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, S., Yan, Z., Guo, Y., Su, X., Cao, Y., Jiang, B. (2021). SPM-IS: an auto-algorithm to acquire a mature soybean phenotype based on instance segmentation. *Crop J.* doi: 10.1016/j.cj.05.014. [Epub ahead of print].

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017a). "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (San Juan, PR), 2117–2125.

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017b). "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 2980–2988. doi: 10.1109/ICCV.2017.324

Liu, G., Nouaze, J. C., Touko Mbouembe, P. L., and Kim, J. H. (2020). YOLO-tomato: a robust algorithm for tomato detection based on YOLOv3. *Sensors* 20:2145. doi: 10.3390/s20072145

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., and Fu, C. (2016). "Ssd: single shot multibox detector," in *European Conference on Computer Vision* (Cham: Springer), 21–37. doi: 10.1007/978-3-319-46448-0_2

Maini, R., and Aggarwal, H. (2010). A comprehensive review of image enhancement techniques. *arXiv preprint arXiv:4053*.

Mao, S., Ma, L., i,. Y., Zhang, Y., Zhou, B., Wang, J., et al. (2020). cucumber recognition algorithm for harvesting robots in the natural environment using deep learning and multi-feature fusion. *Comput. Electron. Agric.* 170:105254. doi: 10.1016/j.compag.2020.105254

Qin, X., Huang, R., and Hua, B. (2021). Research and implementation of yield recognition of Citrus reticulata based on target detection. *J. Phys. Conf. Ser.* 1820:012128. doi: 10.1088/1742-6596/1820/1/012128

Rai, R. K., Asnani, J., and Sontakke, T. R. (2012). Review of shrinkage techniques for image denoising. *Int. J. Comput. Appl.* 42, 13–16. doi: 10.5120/5799-8009

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* (Las Vegas, NV), 779–788.

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Honolulu, HI), 7263–7271.

Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. *arXiv preprint arXiv*:02767.

Ren, S., He, K., Girshick, R., and Sun, J. (2016). 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

Sharifi, A. (2021). Yield prediction with machine learning algorithms and satellite images. *J. Sci. Food Agric.* 101, 891–896. doi: 10.1002/jsfa.10696

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1556*.

Specht, D. F. (1991). A general regression neural network. *IEEE Trans. Neural Netw.* 2, 568–576. doi: 10.1109/72.97934

Sun, S., Paterson, L. I. C., Chee, A. H., Robertson, P. W., and Image, J. S. (2019). processing algorithms for infield single cotton boll counting and yield prediction. *Comput. Electron. Agric.* 166:104976. doi: 10.1016/j.compag.2019.104976

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., and Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* 157, 417–426. doi: 10.1016/j.compag.2019.01.012

Yang, W., Nigon, T., Hao, Z., Paiao, G. D., Fernández, F. G., and Mulla, D. (2021). Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput. Electron. Agric.* 184:106092. doi: 10.1016/j.compag.2021.106092

Yang, Y., Huang, X., Cao, L., Chen, L., and Huang, K. (2019). "Field wheat ears count based on YOLOv3," in *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)* (Dublin: IEEE), 444–448.

Yu, H., He, F., Pan, Y. (2020). A scalable region-based level set method using adaptive bilateral filter for noisy image segmentation. *Multimedia Tools Appl.* 79, 5743–5765. doi: 10.1007/s11042-019-08493-1

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

Zannou, J. G. N., and Houndji, V. R. S. (2019). "Using machine learning," in *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)* (Paris: IEEE), 1–4.

Zhang, J., Mungara, P., and Jane, J. (2001). Mechanical and thermal properties of extruded soy protein sheets. *Polymer* 42, 2569–2578. doi: 10.1016/S0032-3861(00)00624-8

Zhao, S., Zheng, H., Chi, M., Chai, X., and Liu, Y. (2019). Rapid yield prediction in paddy fields based on 2D image modelling of rice panicles. *Comput. Electron. Agric.* 162, 759–766. doi: 10.1016/j.compag.2019.05.020

Zhao, T., Wang, Z., Yang, Q., and Chen, Y. (2017). Melon yield prediction using small unmanned aerial vehicles[C]//Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II. *Int. Soc. Optics Photon.* 10218:1021808. doi: 10.1117/12.2262412

Zhu, J., Cen, H., He, L., and He, Y. (2019). Development and performance evaluation of a multi-rotor unmanned aircraft system for agricultural monitoring. *Smart Agric.* 1:43. doi: 10.12133/j.smartag.2019.1.1.201812-SA011

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fpls.2022.814747

Optimized Design, Monitoring System Development and Experiment for a Long-Belt Finger-Clip Precision Corn Seed Metering Device

*Han Tang†, Changsu Xu†, Ziming Wang, Qi Wang and Jinwu Wang**

College of Engineering, Northeast Agricultural University, Harbin, China

OPEN ACCESS

Edited by:
Daobilige Su,
China Agricultural University, China

Reviewed by:
Zhenguo Zhang,
Xinjiang Agricultural University, China
Guoming Li,
Iowa State University, United States
Yingjun Pu,
Southwest University, China
Zhang Guozhong,
Huazhong Agricultural University, China

***Correspondence:**
Jinwu Wang
jinwuw@neau.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *14 November 2021*
Accepted: *07 January 2022*
Published: *28 January 2022*

Citation:
Tang H, Xu C, Wang Z, Wang Q and Wang J (2022) Optimized Design, Monitoring System Development and Experiment for a Long-Belt Finger-Clip Precision Corn Seed Metering Device. Front. Plant Sci. 13:814747. doi: 10.3389/fpls.2022.814747

To solve multiple problems, such as the poor seeding process stability in the conventional finger-clip precision corn seed metering device and the inability to monitor the seeding effect, a long-belt finger-clip precision seed metering device was optimized and designed. The overall structure and working principle were described, and the mechanism of smooth transport and delivery was analyzed. A diffuse reflection photoelectric sensor and rectangular optical fiber sensor were used to monitor the number of corn seeds in the seeding process, and the states of multiple and miss seeding were calculated. A corn seeding quality monitoring system was designed. In this study, the seed metering performance of the long-belt finger-clip precision seed metering device was compared to that of the conventional finger-clip precision corn seed metering device. It was shown that the reseeding index, the miss-seeding index and the coefficient of variation can be effectively reduced with increasing seed metering tray speed. At the maximum speed of 65r/min, the qualified index increased from 75.75 to 84.70%, the reseeding index decreased from 13.66 to 8.49%, the miss-seeding index decreased from 10.59 to 6.81%, and the coefficient of variation decreased from 20.69 to 6.83%. The variations of these four evaluation parameters with the seed metering tray rotating speed were analyzed. Furthermore, the effects of the seeding frequency and seeding speed on the four evaluation parameters were studied through single factor and variance analyses. The results showed that the relative errors of the qualified index, the reseeding index, the miss-seeding index and the seeding amount increased gradually with the increase in the seed metering tray rotating speed, and the monitoring accuracy of the sensor decreased gradually. The accuracy of sensor monitoring decreased with increasing seeding frequency and seeding speed. This study provides an optimized scheme for the smooth delivery and movement of conventional seed metering devices and provides a technical reference for the development and design of monitoring systems with multiple index and the miss-seeding index of seed metering devices.

Keywords: seed metering device, optimization design, monitoring system, performance of precision, experiment

INTRODUCTION

Precision seeding has the advantages of conserving seeds, reducing labor intensity and improving operation efficiency, comprising one of the important trends for agricultural development in the future (Yazgi and Degirmencioglu, 2014; Xing et al., 2020). As the core working component, precision seed metering devices are mainly divided into mechanical and pneumatic types (Navid et al., 2011; Miller et al., 2012). The pneumatic seed metering device has the characteristics of good universality and high working speed, but it needs additional air source, has complex structure and high failure rate. In addition, seeds need to be strictly graded. Compared with pneumatic seed metering device, mechanical seed metering device is widely used because of its simple structure, convenient maintenance, low cost, no pressure air source and no sealing. At present, the common mechanical seed metering devices mainly include socket-wheel type, disk type, spoon type and finger-clip type. In the process of corn sowing, finger-clip seed metering device has the characteristics of low damage rate and good adaptability. It is the most widely used among the types of mechanical seed metering device (Hossen et al., 2014; Haque et al., 2016; Soyoye, 2020). The research on finger clip precision corn seed metering device is of great significance to popularize precision sowing technology. However, the problem of poor stability in the transportation and delivery process, caused by the violent vibration of working components, is a common problem in mechanical precision seed metering devices (Zhao et al., 2020). At the same time, the seed metering device process is closed, so it is impossible to directly observe the effect of seeding. If such problems in seed metering devices are not found in time during a large-scale seeder operation process, periodic large-area multiple seeding and miss seeding will occur because of the fast working speed and the wide seeding range (Wang et al., 2020). Therefore, it is necessary to monitor the working quality of the seed metering device to ensure reliability and stability in the process of precision seeding.

The monitoring of the seeding quality is the basis for not only the automation and intelligence of the seeding link but also the reliable guarantee of signal feedback and equipment research along with the development of the follow-up seeding link. A large number of scholars have devoted themselves to seeding quality monitoring research, which includes high-speed photography, piezoelectric sensing, and photoelectric monitoring (Zhao and Jian, 2005; Chen et al., 2009; Li et al., 2010). Among these methods, the high-speed photography method has the advantage of high monitoring accuracy. Karayel et al. (2006) used high-speed camera and image processing technologies to obtain the seed distance and seed distribution uniformity, then judged the qualification of the seed metering device based on the seed distance and uniformity. However, this method involves a large amount of image data and expensive equipment, so it is difficult to popularize and apply in the actual working process. The piezoelectric sensing method identifies the seeding amount by monitoring the change in pulse voltage produced by the seed hitting a piezoelectric film or piezoelectric ceramic (Zhang et al., 2011). Wang et al. (2019) developed an impact seeding amount monitoring sensor for rice hill direct seeding; this sensor can effectively identify the seeding amount of rice seeds hitting the piezoelectric film. Ding et al. (2017) developed a small seed size seeding amount sensor based on the piezoelectric impact principle; this sensor realized the real-time monitoring of the seed metering frequency and total seed metering amount. Huang et al. (2013) developed a precision corn seeding monitoring system based on a polyvinylidene fluoride piezoelectric film, and the monitoring accuracy of miss-seeding rate was 96%. The technical difficulty of the piezoelectric sensing method is related to the selection of appropriate sensitive materials according to the physical characteristics of different seeds, where the installation positions of sensitive materials have higher requirements. Compared with the high-speed photography method and the piezoelectric sensor method, the photoelectric monitoring method has the advantages of a simple sensor structure, fast response speed, and easy development and configuration; therefore, this method has gradually become the focus of research (Al-Mallahi and Kataoka, 2016; Besharati et al., 2019). Karimi et al. (2017, 2019) built a seeding monitoring device and a seeding monitoring system with infrared laser diode array sensors and realized the measurement of seed flow. However, the array can monitor only the seed information in a single direction, easily producing monitoring errors. Kumar and Raheman (2018) arranged an infrared light-emitting diode (LED) as a ring. Such infrared light covered the whole section of the seeding tube, effectively improving the monitoring accuracy and overall accuracy. Due to the different delivery and transport modes of different seed metering devices, a diversity in the seed falling posture often appears in the seeding process (Liu et al., 2018). One of the most common situations is the overlap in the process of seed falling, which will lead to the sensor unable to effectively monitor the number of seeds and the problem of low monitoring accuracy.

Limited by the characteristics and performance of the sensor, the use of a single sensor can meet only the seeding conditions of small flow and low speed. With the increase in the speed of the seed metering tray, a single sensor cannot achieve accurate and effective monitoring. Zhang et al. (2013) installed LEDs and photosensitive resistors in the seed guide tube of a soybean seeder to detect the seeding condition based on the blocked light when the seeds fell. Lan et al. (1999) used an LED with a diameter of 3 mm and a phototransistor photosensor to measure the spacing of small seeds from the seed metering device. Qiu et al. (2019) designed an operation quality monitoring system of a small particle size seeder based on a photoelectric sensor for rectangular infrared detection and a complementary metal-oxide semiconductor (CMOS) image sensor. Visualization of this seeding operation monitoring was realized by seed image acquisition and a photoelectric sensor. Technology based on multisensor information fusion can greatly improve the effective monitoring of seeding quality under multiple working conditions, but the installation positions, configuration forms and monitoring objects of multisensors are not universal. The problem of multisensor configuration and development for specific crop monitoring needs to be solved.

To solve problems such as the poor seeding process stability in the conventional finger-clip precision seed metering device used

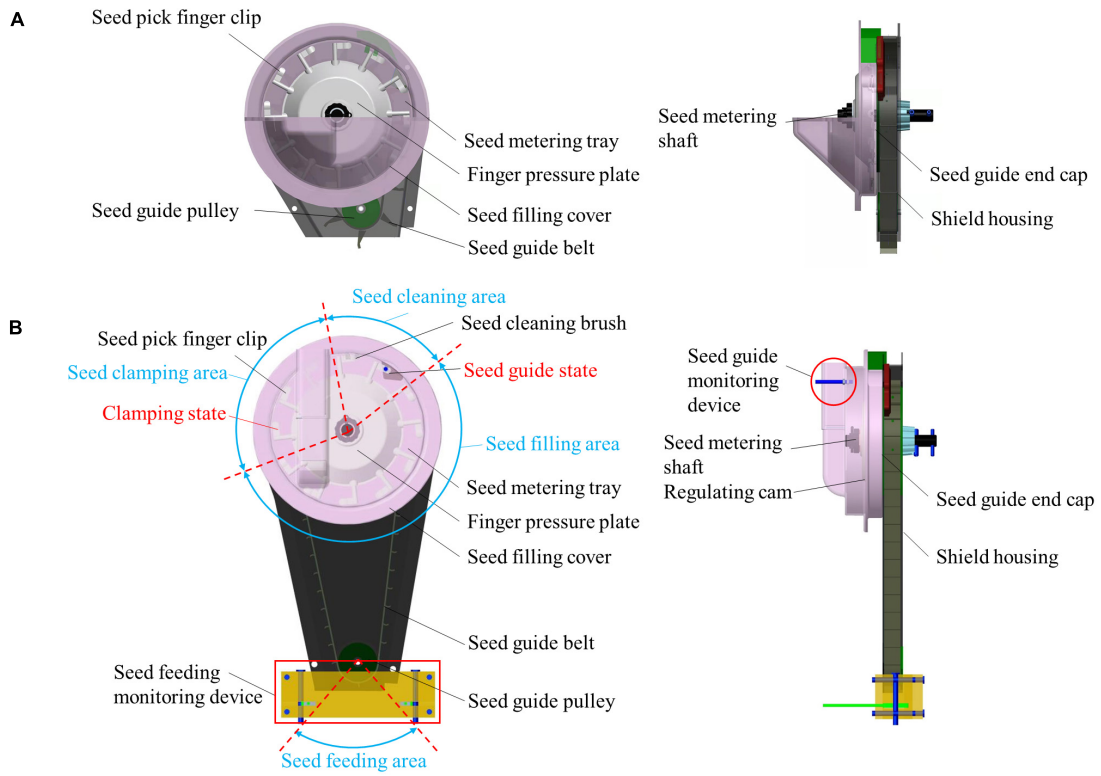


FIGURE 1 | Overall structure. **(A)** Overall structure of conventional finger-clip precision corn seed metering device. **(B)** Overall structure of long-belt finger-clip precision corn seed metering device.

for corn and the inability to monitor the seeding effect, the long-belt finger-clip precision seed metering device used for corn was optimized and designed. The mechanism of smooth transport and delivery was analyzed. A diffuse reflection photoelectric sensor and a rectangular optical fiber sensor were used to monitor the seeding state information. Based on these sensors, a quality monitoring system for precision corn seed metering devices was developed. The effectiveness of the long-belt finger-clip precision corn seed metering device for smooth migration and the reliability of the monitoring system were verified by comparative experiments, single factor experiments and variance analysis. This study provides an optimized scheme for smooth delivery and movement in conventional seed metering devices and provides a technical reference for monitoring system development and design with multiple index and the miss-seeding index of seed metering devices.

MATERIALS AND METHODS

Overall Structure and Working Principle of the Long-Belt Finger-Clip Precision Corn Seed Metering Device and Monitoring System

Conventional finger-clip precision corn seed metering device is mainly composed of a seed pick finger clip, seed guide pulley, seed metering tray, finger pressure plate, seed filling cover, seed guide belt, seed metering shaft, seed guide end cap, shield housing. The overall structure is shown in **Figure 1A**. The conventional finger-clip precision corn seed metering device is equipped with a seed guide tube under it, so it is unable to sowing at zero speed, resulting in the reduction of seed metering performance when the seeds bounce and collide. In addition, the structural configuration cannot meet the layout requirements of sensors. In this study, the seed guide tube structure was replaced by the lengthened seed guide belt, which simplified the overall structure. At the same time, the research was carried out from the aspect of stable delivery mechanism to achieve the effect of zero speed seed delivery. A long-belt finger-clip precision corn seed metering device is designed. It is mainly composed of a seed guide pulley, seed guide belt, seed filling cover, finger pressure plate, seed metering tray, seed cleaning brush, seed-pick finger-clip, seed feeding monitoring device, shield housing, seed guide end cap, seed guide monitoring device, regulating cam and seed metering shaft. The overall structure is shown in **Figure 1B**. The parameters of long-belt finger-clip precision corn seed metering device is shown in **Table 1**. Among these components, the finger pressure plate and the regulating cam are two of the key working components in the seed metering device, and the rationality of the structure configuration directly affects the working quality of the seed metering device. The finger pressure plate is a combination of 12 finger clips connected by

TABLE 1 | Parameters of long-belt finger-clip precision corn seed metering device.

Item	Value	Unit
Number of finger clips	0–12	/
Diameter of seed guide pulley	50	mm
Diameter of seed guide pulley I	200	mm
Seed guide belt spacing	40	mm
Seed metering disk speed	0–65	r/min

fine-tuning springs (the number of finger clips can be adjusted freely, and up to 12). The finger pressure plate and regulating cam are installed on the inside of the finger clamping plate in turn. The seed metering tray is made of galvanized steel to increase the friction characteristics of the corn seeds. The seed guide belt is made of rubber, and its circumference is equipped with 12 inclined blades. The 12 seed guide rooms are formed with the seed guide end cap and the machine shell. The seed cleaning brush is made of bristles, and the angle of the brush can be artificially adjusted to control the seed cleaning degree. The seed guide monitoring device is situated at the seed guide port, and a diffuse reflection photoelectric sensor is used to monitor the number of theoretical seeds (the number of seed-pick finger-clips is monitored through the diffuse reflection photoelectric sensor, were a seed-pick finger-clip carries a corn seed in theory). The seed feeding monitoring device is arranged under the seed feeding port, and a rectangular optical fiber sensor is used to monitor the number of seed guide belts delivered to the soil.

The working area of the seed metering device can be divided into four stages: the seed filling area, the seed holding area, the seed cleaning area and the seed feeding area. During operation, the seed filling room is filled with the corn seeds from the seed box. The walking wheel of the machine transmits the power to the seed metering shaft through chain transmission and drives the finger pressure plate and the seed-pick finger-clip to rotate. The fixed cam and the fine-tuning spring work together to control the timing of the finger clip opening and closing. When the seed-pick finger-clip is opened, the seed is filled and clamped in the seed filling area. When the seed-pick finger-clip is closed, multiple seeds are withheld from the seed filling area to complete the seed filling process. The seed-pick finger-clip rotates and moves smoothly in the seed holding area. When the seed is moved to the seed cleaning area, the seed cleaning brush removes the multiple seeds and completes the seed cleaning process. A single grain of corn is placed in the seed guide room on the back of the seed metering device from the seed guide mouth, and the seed guide belt rotates synchronously with the finger pressure plate. A single seed is smoothly deposited into the soil to complete the process of seeding, and precision seeding is achieved. To realize the quality monitoring of a precision seed metering device, a diffuse reflection photoelectric sensor and a rectangular optical fiber sensor are set up under the seed guide port and the seed feeding port, respectively. When the seed is transported and delivered, the sensor is shielded to return a low-level signal. According to the pulse signal returns by the sensor and processes by the single-chip microcomputer, the seed falling time interval is counted and compared to the set theoretical time interval. The reseeding index and miss-seeding index in the process of precision seeding are calculated.

Analysis of the Mechanism for Stable Transport and Delivery

The lengthened seed guide belt of the long-belt finger-clip precision corn seed metering device can reduce the delivery height of the seed to the soil and slow down the bouncing collision frequency between the seed and the soil in the process of high-speed precision seeding. The seeding accuracy, uniformity and horizontal and vertical straightness are improved. In the seeding process, the seed metering shaft drives the seed metering tray to rotate, and the seed-pick finger-clip rotates the single seed counterclockwise; the single seed is then smoothly transported to the seed guide mouth for seeding. The first migration operation is realized. The seed is thrown onto the seed guide belt, which rotates synchronously with the seed metering tray through the seed guide mouth and moves smoothly to the seed delivery point under the action of gravity and the supporting force. The zero speed in low position of the seed to the soil is realized. The second migration operation is realized. The relative speed of the seeds falling into the seed ditch is offset by secondary delivery, and smooth seed transport and delivery is realized, as shown in **Figure 2**.

To study the stationarity of the seeds in the transport and delivery stage, the motion state of the seeds is mechanically analyzed and the critical conditions of the relative balance between the seeds and the seed guide blade are studied (Wang et al., 2017). Taking the rotating center of the seed metering shaft as the coordinate origin O_1, the spatial Cartesian coordinate system XYZ is established, as shown in **Figure 3**. When a single seed enters the seed guide room through the seed guide port, the force of the seed is analyzed when it rotates counterclockwise with

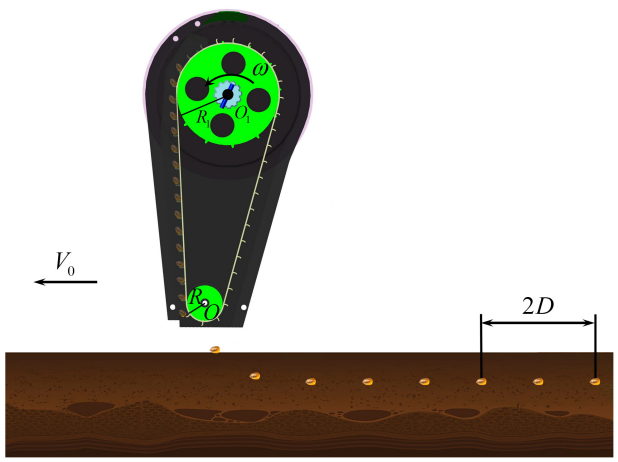


FIGURE 2 | Schematic diagram of smooth transport and delivery. D is the corn seed spacing, mm; V_0 is the working speed of the machines, mm/s; ω is the angular velocity of the seed metering shaft, rad/s.

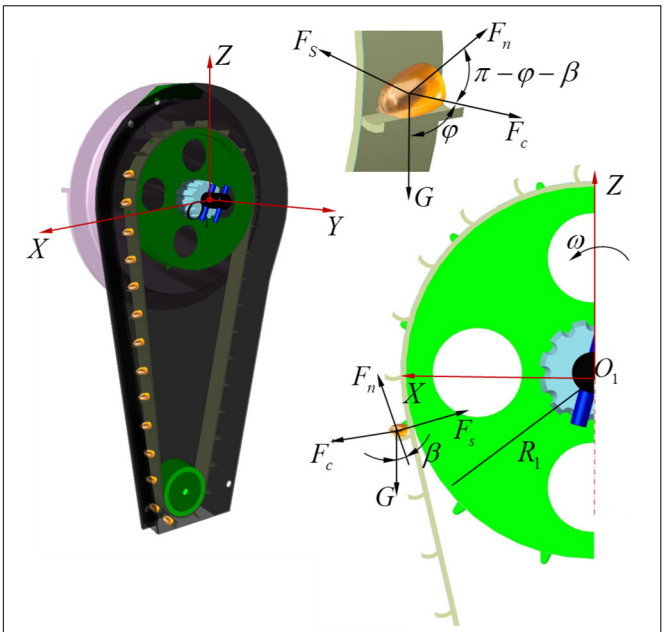


FIGURE 3 | Force analysis diagram of the stable transport and delivery process.

the seed guide belt. The seeds are affected by the spatial force system composed of the centrifugal force F_c, the supporting force of the blade F_n, the friction force of the blade F_s and the seed gravity G.

If the relative balance between the seed and the seed guide blade is ensured and the seed is not thrown away from the seed guide blade, the force along the slip direction of the seed relative to the blade should be satisfied in the rotating plane XO_1Z of the seed guide belt.

$$F_c + G\cos\phi + F_n\cos(\pi - \beta - \phi) \le F_s\cos\beta \tag{1}$$

Among these values,

$$\begin{cases} F_c = m\omega^2 R_1 \\ F_n = mg\sin\beta \\ F_s = \mu F_n \end{cases} \tag{2}$$

where m is the mass of the corn seed, mg; ϕ is the relative rotation angle of the seed guide blade, (°); β is the structural inclination angle of the seed guide blade, (°); μ is the friction coefficient between the corn seed and the seed guide blade, set to 0.15; R_1 is the radius of the seed guide pulley I, mm; and F_c is the centrifugal force of the corn seeds, N.

Equation (3) is obtained by the combination of Equation (1) and Equation (2).

$$m\omega^2 R_1 + mg\cos\phi \le mg\sin\beta(\mu\cos\beta + \cos(\beta + \phi)) \tag{3}$$

To solve the limiting value of the seed metering speed under the critical slip state, Equation (3) is arranged as follows:

$$\omega \le \sqrt{\frac{g\sin\beta(\mu\cos\beta + \cos(\beta + \phi)) - g\cos\phi}{R_1}} \tag{4}$$

When the speed of the seed metering tray satisfies Equation (4), the corn seed and the seed guide belt remain relatively static and relative slip does not occur. Equation (4) shows that the factors affecting the speed limit of the smooth seed movement and delivery are related to the relative rotation angle of the seed guide blade, the structural inclination angle of the seed guide blade and the radius of the seed guide pulley I. Moreover, this limit is related to the friction coefficient between the corn seeds and the seed guide blade but has nothing to do with the mass of the corn seeds.

Design of the Seeding Monitoring System

The hardware system and software system of the seeding monitoring system were designed to accurately monitor the important seed metering indexes, such as the qualified index, reseeding index and miss-seeding index, in the metering process of a long-belt finger-clip precision corn seed metering device.

Hardware System Design

The hardware system is mainly composed of a STM32F103 single-chip microcomputer, universal serial bus (USB) to transistor-transistor logic (TTL) chip, upper computer, integrated operational amplifier, diffuse reflection photoelectric sensor, rectangular grating sensor and fixed device, as shown in **Figure 4**. Each part is electrically connected through the communication lines, the signal lines and the power lines to complete the information exchange. Among these components, the power lines are used for the power supply and connection between each piece of equipment, where the power supply is the direct current (DC) 12 V vehicle battery. The communication line adopts the RS232 serial communication network, as designed according to the communication protocol, which was beneficial to the expansion of the system. A rectangular grating sensor adopts the principle of infrared radiation, which changes the incident light path in the seeding state and changes the light intensity received by the infrared receiver. The signal is processed into a pulse signal by the signal amplifier. The single-chip timer (time 4, time 5) is configured to input capture mode, and the corresponding input/output (IO) port is set as the input mode. Thus, the pulse signal was obtained, and the planting parameters were calculated. At the same time, this system cooperated with the diffuse reflection photoelectric sensor installed on the seed guide port, the times of the seed-pick finger-clip passing through the seed guide port were monitored and the rotating speed was calculated according to the reflection principle of light. Based on the times and rotating speed of the seed-pick finger-clip monitored by the diffuse reflection photoelectric sensor, multiple seeding occurred when the data monitored by the rectangular grating sensor were larger than those monitored by the diffuse reflection photoelectric sensor.

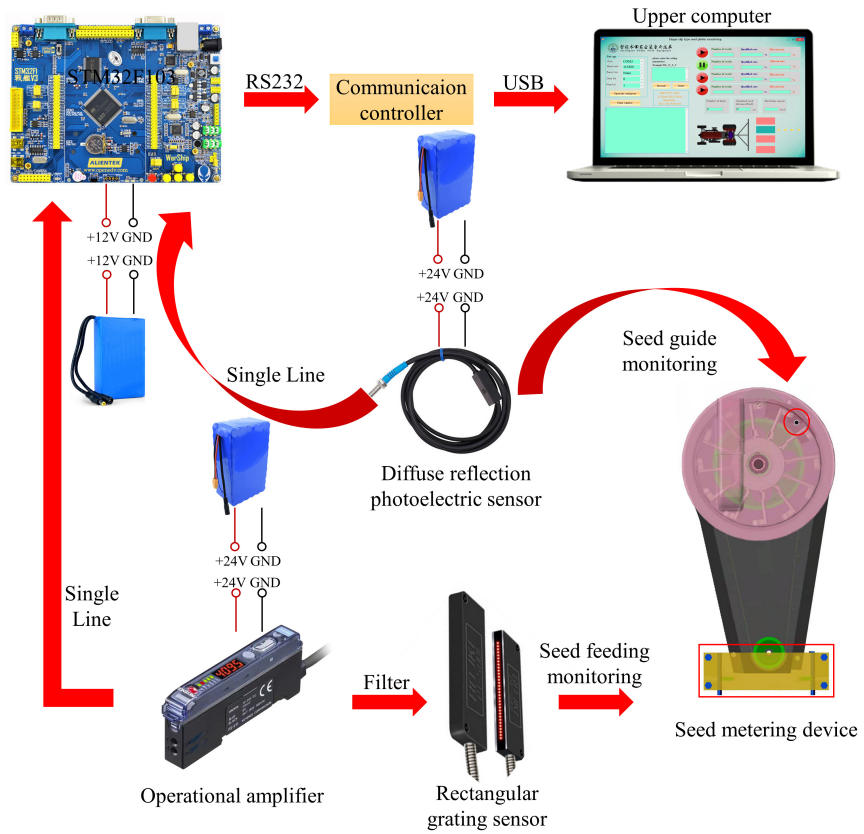


FIGURE 4 | Hardware system.

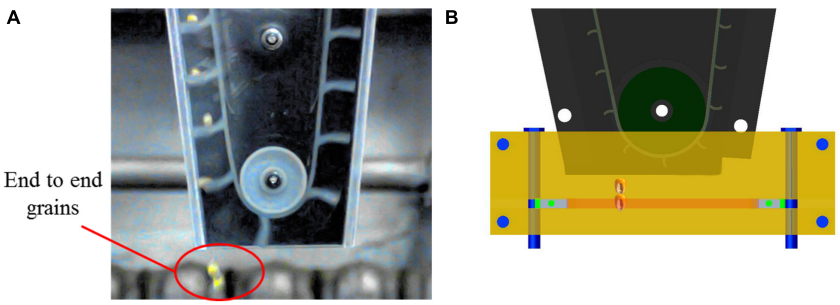


FIGURE 5 | Schematic diagram of multiple seeding. **(A)** Rseeding in high-speed camera observation. **(B)** Principle of reseeding monitoring.

When the data monitored by the rectangular grating sensor were smaller than those monitored by the diffuse reflection photoelectric sensor, miss seeding occurred. The data detected by the sensor were transmitted to the STM32F103 single-chip microcomputer through the signal line for control and calculation, and the serial port signal was collected by the upper computer through the USB to TTL chip to obtain the seeding information.

Software System Design

Seed metering monitoring was mainly realized by monitoring the number of input seeds, the number of rotations and the rotating speed of the seed-pick finger-clip through the seed guide port. Then, a diffuse reflection photoelectric sensor and a rectangular grating sensor converted this information into a pulse signal, and the system counted the number of seeds and the number of times the seed-pick finger-clip (theoretical planting number) was placed in the seed guide port by monitoring the falling edge of the pulse signal. Timing was determined by monitoring the rising edge and falling edge of a pair of pulse signals. The seeding monitoring sensor time was recorded as T_A. The interval time of the seed guide monitoring sensor was recorded as T_B.

According to the high-speed camera experiment, the rectangular grating sensor received two pairs of pulse signals in

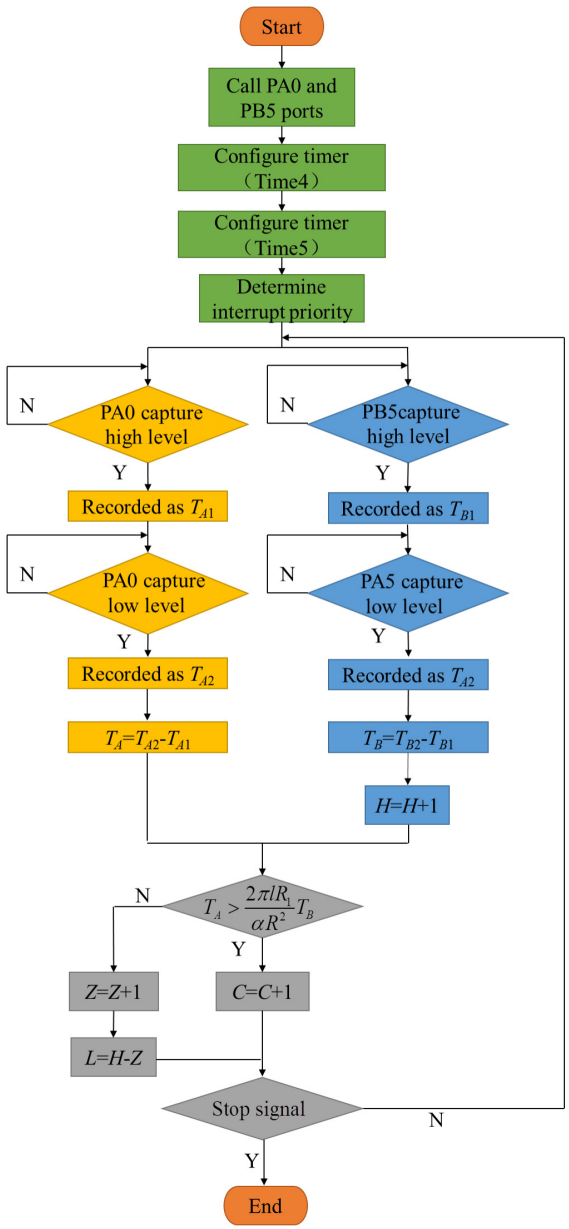


FIGURE 6 | Program flow chart. *L* is the miss-seeding amount; *C* is the reseeding amount; *H* is the theoretical seeding amount; and *Z* is the total amount of seeding.

succession when multiple seeding events occurred. However, the diffuse reflection photoelectric sensor received a pair of pulse signals only once, and the number of multiple seedings was increased by one. The end-to-end connection occurred in the vertical plane during the fall of the multiple seeding, as shown in **Figure 5**. The interval time of the rectangular grating sensor T_A increased, which resulted in the omission of multiple seeding counts. Therefore, the number of multiple seedings was increased by one when Equation (5) was satisfied. The system program flow chart is shown in **Figure 6**. Firstly, enable the logic pins of single chip PA0 and PB5, and the corresponding timers (Time4 and Time5) are configured to enable the single chip microcomputer to receive external signals and have timing ability. In order to determine the sequence of receiving or transmitting signals, the priority of receiving interrupt and transmitting interrupt is set. At this time, PA0 and PB5 wait for the external signal to trigger. When any pin voltage becomes high, the timer register stores the current time T_{A1} or T_{B1}. When the external signal disappears, the pin voltage becomes low, and the timer register stores the current time T_{A2} or T_{B2}. The CPU calculates the signal receiving time T_A or T_B, and the theoretical sowing amount is increased by one grain. The CPU determines whether there is a multiple seeding amount and miss seeding amount according to the discrimination relationship. Finally, the CPU determines whether the stop signal is received. If the stop signal is received, the program stops, otherwise it enters the next cycle.

$$T_A > \frac{2\pi lR_1}{\alpha R^2} T_B \tag{5}$$

where T_A is the total time of the seed occlusion rectangular grating sensor, s; T_B is the total time based on the finger clip occlusion diffuse reflection photoelectric sensor, s; l is the length of the seed, mm, the vernier caliper was used to measure the length of 1,000 "Demeiya No. 1" maize seeds widely planted in Northeast China, and the average value was taken, taking 9.35mm; and α is the rotation angle of the seed-pick finger-clip, (°).

In this study, a Windows-form application program was developed and designed based on the C# language; this program can display the monitoring seeding parameters in real time. Additionally, when the phenomena of multiple seeding and miss seeding occurred, a seeding fault alarm occurred in the software interface. The software interface is shown in **Figure 7**.

According to the seeding function, the software interface can be divided into four areas: the communication string number setting, seed parameter setting, operation parameter monitoring and miss seeding alarm. Among these areas, the communication string number setting area is used for the selection and opening and closing of the working string number for the equipment communication. The parameter setting area is used to design the relevant parameters, such as the length of the finger clip. After the parameter design is completed, it is saved to the running memory, and the corresponding parameters are read when the program is running. The monitoring area of the operation parameters includes the number of seedlings, the theoretical number of seedlings, the interval seeding time and the rotation speed of the seed metering tray. Among these parameters, the number of seedlings is used to display the seeding number of each row for the seed metering device. The theoretical number of seedlings is used to determine the number of seedlings in each row under normal operating conditions (without miss seeding and multiple seedlings). The phenomenon of multiple seeding is determined through the rotational speed of the seed metering tray and the seeding time interval. When miss seeding occurs continuously, the miss seeding alarm displays a flashing red light to inform the user. Through the abovementioned method, the

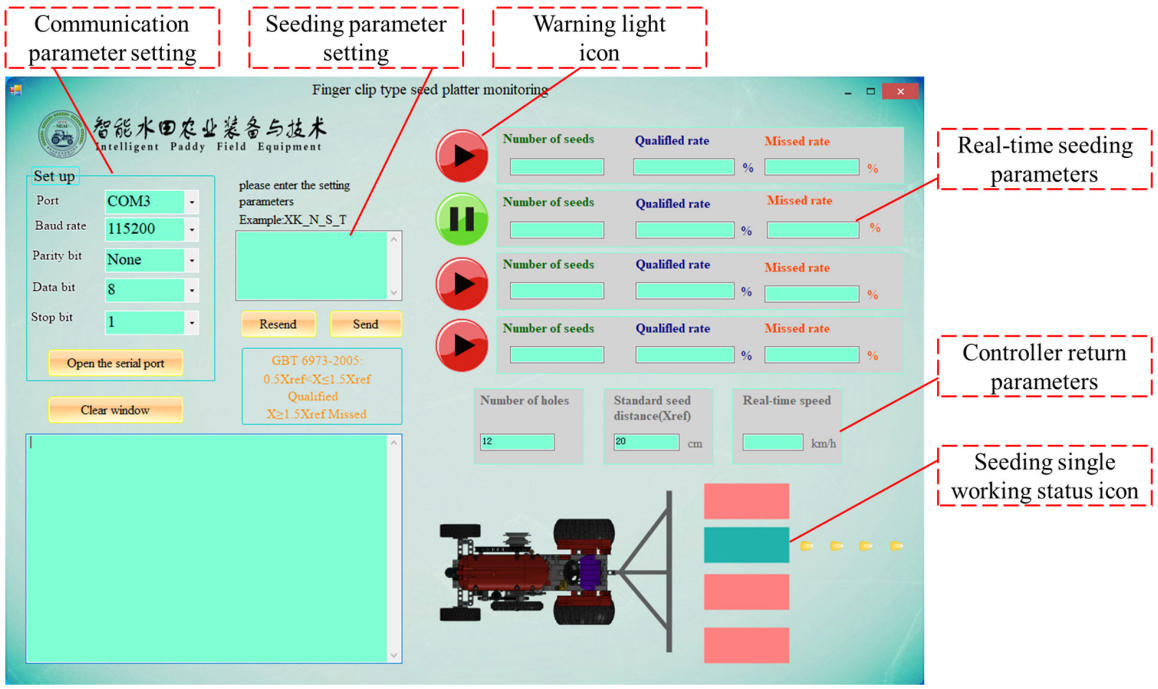


FIGURE 7 | Software interface.

online monitoring of the qualified index, the reseeding index and the miss index can be realized.

Experiment

To compare the seeding performances of the long-belt finger-clip precision corn seed metering device and conventional finger-clip precision corn seed metering device and to detect the accuracy of the designed monitoring system, a bench experiment was carried out on seeding performance. The experimental site was the Seeding Performance Laboratory of Northeast Agricultural University. Demeya No. 1 corn seed was selected as the experimental variety. The corresponding 1,000-grain mass was 281.12 g, the average density was 1.154 g/cm^3, and the average length was 9.27 mm, width was 7.40 mm, and thickness was 4.11 mm (the average values were measured based on 100 seeds). The experimental device was the JPS-12 seed metering device performance experiment bench developed by the Heilongjiang Agricultural Machinery Engineering Research Institute. According to GB/T6973-2005 "Experiment method of single seed (Precision) seeder," the experiment factors were selected as the rotating speed of seed metering tray, which were 15 r/min, 25 r/min, 35 r/min, 45 r/min, 55 r/min, and 65 r/min, respectively. The experimental indexes included the qualified index, the reseeding index, the miss-seeding index and the coefficient of variation to evaluate the seeding stability of the seed metering device. Additionally, the accuracy of the monitoring sensor was evaluated based on the relative deviation of the qualified index, the reseeding index, the miss-seeding index and seeding amount. The greater the relative deviation, the lower the monitoring accuracy of the seed metering device. There were 1,000 seeds measured in each group, the experiment was repeated three times, and the results were averaged. During the experiment, the spacing between adjacent corn seeds was measured, and the number of seeds was counted to calculate each index. The calculation equation of each index was as follows:

$$\begin{cases} Q = \dfrac{n_0}{N} \times 100\% \\ M = \dfrac{n_1}{N} \times 100\% \\ E = \dfrac{n_2}{N} \times 100\% \\ V = \sqrt{\dfrac{\sum (x - \bar{x})}{(n' - 1)\,\bar{x}^2}} \times 100\% \\ Q' = \dfrac{\left| n_0^{'} - n_0 \right|}{n_0} \times 100\% \\ M' = \dfrac{\left| n_1^{'} - n_1 \right|}{n_1} \times 100\% \\ E' = \dfrac{\left| n_2^{'} - n_2 \right|}{n_2} \times 100\% \\ S = \dfrac{n_a - n_b}{n_a} \times 100\% \end{cases} \tag{6}$$

where Q is the qualified index, %; M is the reseeding index, %; E is the miss-seeding index, %; V is the coefficient of variation, %; Q' is the relative error of the qualified index, %; M' is the relative

error of the reseeding index, %; E' is the relative error of the miss-seeding index, %; n_0 is the manual hole count in one seed; n_1 is the manual hole count for more than one seed; n_2 is the manual hole count without seeds; N is the manual hole count of the total number of holes; $n_0^{'}$ is the monitoring result for the holes in one seed; $n_1^{'}$ is the monitoring result for the holes in more than one seed; $n_2^{'}$ is the monitoring result for holes without seeds; $n^{'}$ is the total number of sample-seeds spacings; x is the theoretical seed-spacing of seeding, mm; $\overline{x}$ is the average distance between sample points, mm; S is the relative error of the seeding numbers; n_a is the result of manually counting the total number of holes; and n_b is the monitoring result for the total number of holes.

The seed metering device was fixed on the mounting frame in the experimental process. The bed belt moved in the opposite direction relative to the seed metering device. The forward motion state of the seeder was simulated. The fuel injection pump sprayed oil on the bed belt. The corn seed fell from the seeding mouth to the seed bed belt. Real-time detection and data acquisition were carried out through the camera processing device to accurately measure the seeding performance index, as shown in **Figure 8**.

In May 2021, a field validation experiment was carried out in Acheng District, Harbin City, Heilongjiang Province (127.30 East longitude and 45.33 North latitude). Field environment: straw coverage is (2.08 - 3.16) kg/m^2, straw stubble height is (130 - 150) mm, soil moisture content is 18 - 22%, and soil firmness is (16.05 - 20.41) MPa. Based on the long-belt finger-clip precision corn seed metering device and the developed monitoring system, the no tillage sowing device was integrated, as shown in **Figure 8B**. The power unit is John Deere 454 tractor. During the test, the rotation speed of the seed metering plate shall be consistent with the bench test, so the forward speed of the tractor shall be calibrated before the test. When the forward speed is 3, 5, 7, 9, 11, 13 km/h, the corresponding seed metering disk speed is 15, 25, 35, 45, 55, 65 r/min. The seed spacing is 220 mm. After sowing, the soil layer shall be removed manually to determine the spacing of corn seeds, and errors caused by human factors shall be avoided as far as possible. Each group of experiments was repeated 3 times, and the results were taken as the average value.

The influence of the seed metering tray rotating speed on the sensor monitoring accuracy can be comprehensively analyzed, in essence, as the influence of the sensed frequency and speed of the seed on the sensor monitoring accuracy (Maleki et al., 2006). As the seed metering tray rotated coaxially with seed guide pulley I, seed guide pulley I and the seed guide pulley rotated through the seed guide belt. The speed and frequency of the seeds passing over the two sensors were the same. To further explore the relationship between these two factors, the corresponding values were controlled by adjusting the number of finger clips and the rotation speed of the seed metering tray.

The sensed seed frequency can be obtained based on the rotating speed of the seed metering tray as follows:

$$f = \frac{1}{5}n \tag{7}$$

where f is the seed frequency over the sensor, seeds/s; and n is the rotating speed of the seed metering tray, r/min.

The seed speed over the sensor was calculated (ignoring the distance between the rectangular optical fiber sensor and the bottom of the seeding guide pulley II) as follows:

$$v = \frac{\pi R_1 n}{30000} \tag{8}$$

where v is the seed speed through the sensor, m/s.

RESULTS

Comparative Analysis of the Seeding Performances Between Long-Belt Finger-Clip and Conventional Finger-Clip Precision Corn Seed Metering Devices

The comparison of the seeding performance between the long-belt finger-clip precision corn metering device and the conventional finger-clip precision corn seed metering device is shown in **Figure 9**. The regression equation in **Figure 9** is used to analyze and predict the relationship between performance and seed metering tray speed. The greater the slope of regression equation, the greater the change of seed metering performance index with factors. R^2 represents the overall fit of the regression equation. The maximum value of R^2 is 1. The greater R^2, the better the fitting degree of the regression equation. **Figure 9** shows that the seeding performances of the long-belt finger-clip precision corn seed metering device and the conventional finger-clip precision corn seed metering device decreased with increasing speed for the seed metering tray. When the rotational speed of the seed metering tray was 15-25 r/min, the slope of the coefficient of variation for the conventional finger-clip precision corn seed metering device is greater than that of the long-belt finger-clip precision corn seed metering device. The results showed that with the increase in the rotational speed of the seed plate, the stability of the conventional finger clip precision corn seed metering device and the consistency between the seed spacing were poor. When the rotational speed of the seed metering tray was 25-65 r/min, the qualified index of the conventional finger-clip precision metering device decreased greatly, from 90.83 to 75.75%. The reseeding index, the miss-seeding index and coefficient of variation increased greatly. The reseeding index increased from 4.52 to 13.66%, the miss-seeding index increased from 4.65 to 10.59%, and the coefficient of variation increased from 7.88 to 20.69%. However, the qualified index, the reseeding index, the miss-seeding index and coefficient of variation in the long-belt finger-clip precision corn seed metering device changed, and the overall change was relatively stable. When the rotation speed of seed metering disk is 65 r/min, the qualification index, reseeding index, miss-seeding index and variation coefficient are 84.70, 8.49, 6.81, and 6.83%, respectively. The results showed that the long-belt finger-clip precision seed metering device can smoothly transport and deliver corn seeds and effectively reduce the reseeding index, miss-seeding index and coefficient of variation compared to the conventional finger-clip precision seed metering device.

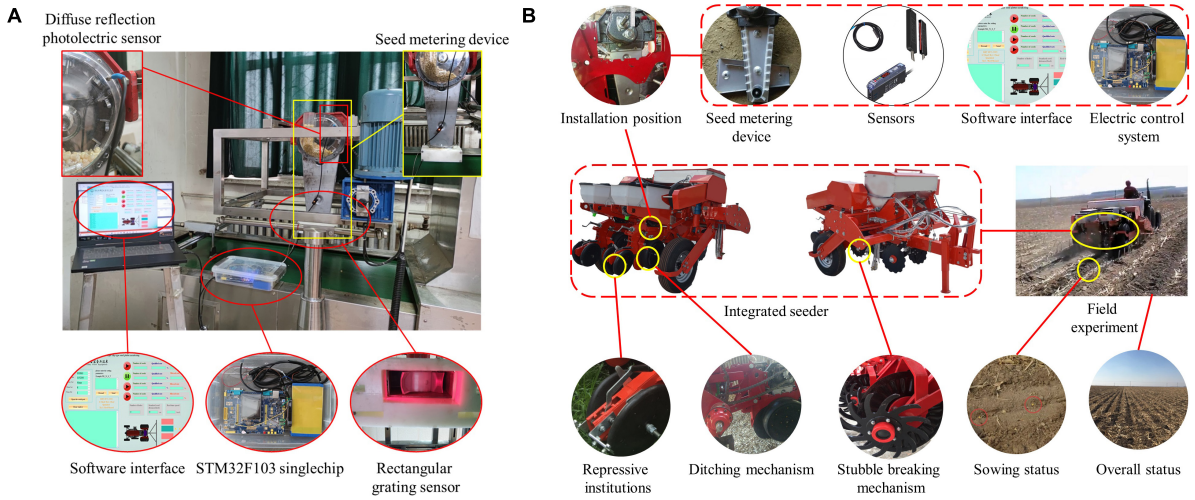


FIGURE 8 | Experiment. **(A)** Bench experiment. **(B)** Field experiment.

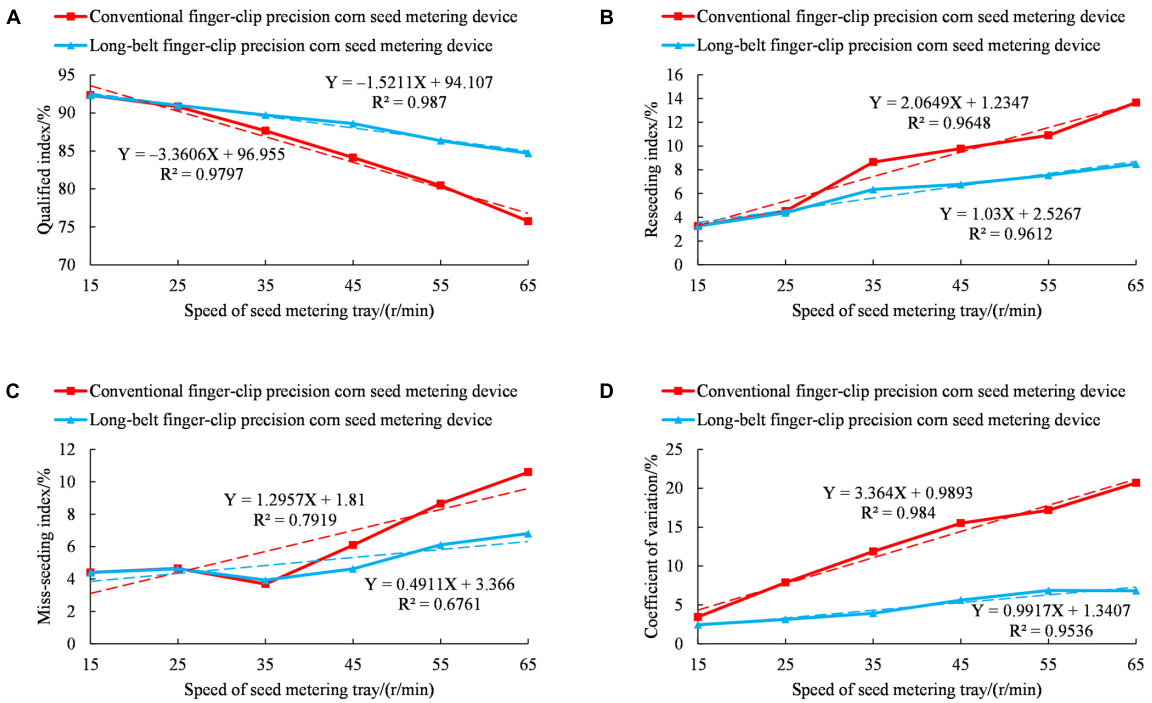


FIGURE 9 | Comparative analysis of the seeding performances for the long-belt finger-clip and conventional finger-clip precision corn seed metering devices. **(A)** Influence of the rotating speed of the seed metering tray on the qualified index. **(B)** Influence of the rotating speed of the seed metering tray on the reseeding index. **(C)** Influence of the rotating speed of the seed metering tray on the miss-seeding index. **(D)** Influence of the rotating speed of the seed metering tray on the coefficient of variation.

Influence of the Rotation Speed of the Seed Metering Tray on the Monitoring Accuracy of the Sensor

To explore the influence of the rotating speed of the seed metering tray on the monitoring accuracy of the sensor, the qualified number, the multiple seeding number, the miss seeding number and the seeding number monitored by the sensor were compared to the data obtained by manual measurement. The variation rules for the relative errors of the qualified index, the reseeding index, the miss-seeding index and seeding amount with the rotating speed of the seed metering tray are shown in **Figure 10**.

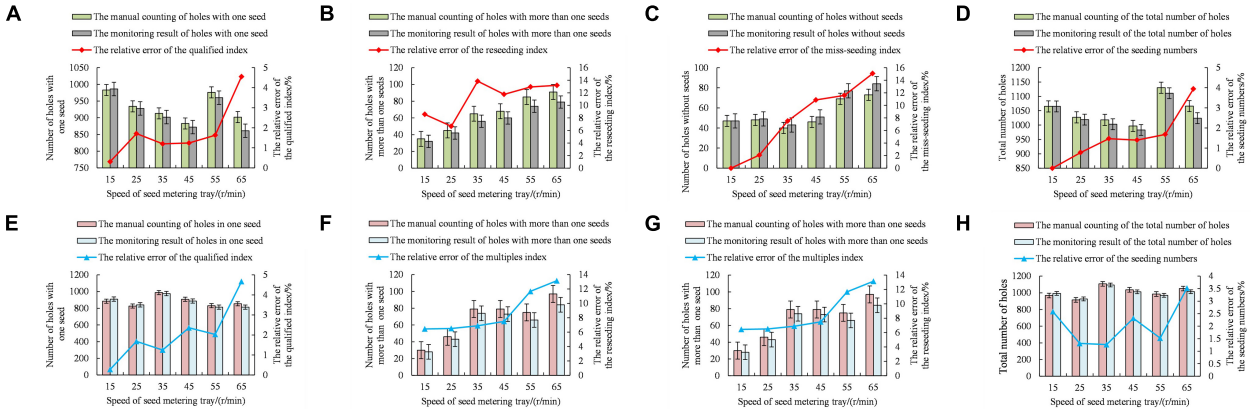


FIGURE 10 | Influence of the rotating speed of the seed metering tray on the monitoring accuracy of the sensor. **(A)** Qualified number in bench experiment. **(B)** Reseeding number in bench experiment. **(C)** Miss-seeding number in bench experiment. **(D)** Seeding amount in bench experiment. **(E)** Qualified number in field experiment. **(F)** Reseeding number in field experiment. **(G)** Miss-seeding number in field experiment. **(H)** Seeding amount in field experiment.

TABLE 2 | Comparison results between bench experiment and field experiment of long-belt finger-clip precision corn seed metering device and monitoring system.

Speed of seed metering tray (r/min)		**15**	**25**	**35**	**45**	**55**	**65**
Performance of long-belt finger-clip precision corn seed metering device	Qualified index in bench experiment (%)	92.34	90.99	89.72	88.6	86.35	84.7
	Qualified index in field experiment (%)	91.46	90.38	89.24	87.65	84.33	81.26
	Reseeding index in bench experiment (%)	3.26	4.38	6.35	6.77	7.54	8.49
	Reseeding index in field experiment (%)	3.11	5.02	7.12	7.65	7.61	9.23
	Miss-seeding index in bench experiment (%)	4.40	4.63	3.93	4.63	6.11	6.81
	Miss-seeding index in field experiment (%)	5.43	4.60	3.64	4.70	8.06	9.51
	Coefficient of variation in bench experiment (%)	2.46	3.15	3.93	5.63	6.87	6.83
	Coefficient of variation in field experiment (%)	2.39	3.27	4.02	5.54	6.96	7.8
Performance of monitoring system	The relative error of qualified index in bench experiment (%)	0.31	1.71	1.2	1.24	1.63	4.55
	The relative error of qualified index in field experiment (%)	0.28	1.68	1.25	2.36	2.03	4.67
	The relative error of reseeding index in bench experiment (%)	8.57	6.67	13.85	11.76	12.94	13.19
	The relative error of reseeding index in field experiment (%)	6.45	6.52	6.89	7.5	11.67	13.15
	The relative error of miss-seeding index in bench experiment (%)	0	2.08	7.5	10.87	11.59	15.07
	The relative error of miss-seeding index in field experiment (%)	3.13	3.03	7.31	7.14	13.46	15.71
	The relative error of seeding numbers in bench experiment (%)	0	0.78	1.47	1.4	1.68	3.94
	The relative error of seeding numbers in field experiment (%)	2.58	1.31	1.26	2.32	1.52	3.52

Figure 10A shows that the relative errors of the qualification index, the reseeding index, the miss-seeding index and seeding amount increased gradually with increasing speed for the seed metering tray in the bench experiment. The results showed that the monitoring accuracy of the sensor decreased with the increasing rotation speed of the seed metering tray. The relative error of the reseeding index and the miss-seeding index was larger than that of the qualified index and seeding amount, mainly due to the small number of multiple seeding holes, and the miss seeding holes were larger as a whole when compared with the manually measured data. However, the overall trend reflected the changing law of the sensor monitoring accuracy.

Figure 10B shows that the relative errors of the qualification index, the reseeding index, the miss-seeding index increased gradually with increasing speed for the seed metering tray in the field experiment. It is verified that the monitoring accuracy of the sensor decreases gradually with the increase of the rotation speed of the seed metering disk. **Table 2** is the comparison results between bench experiment and field experiment of long-belt finger-clip precision corn seed metering device and monitoring system. The performance of the seed metering device in the field experiment is slightly lower than that in the bench experiment, but the data obtained by the monitoring system is less different from that in the bench experiment, and the performance is relatively stable.

The seeding frequency, seeding speed and speed of the seed metering tray are shown in **Table 3**.

To further explore the influences of the seeding frequency and seeding speed on the monitoring accuracy of the sensor, the relative errors of the qualified index, the reseeding index, the miss-seeding index and seeding amount were used as evaluation parameters. Through the value of fixed level 4 (Wang et al., 2017),

TABLE 3 | Seeding frequency and speed corresponding to the rotating speed of the seed metering tray.

Level	Speed of seed metering tray (r/min)	Seeding frequency (seeds/s)	Seeding speed (m/s)
1	15	3	0.09
2	25	5	0.15
3	35	7	0.22
4	45	9	0.28
5	55	11	0.35
6	65	13	0.41

single factor experiments were carried out on the seeding frequency and seeding speed.

First, linear regression analysis of the metering frequency was carried out. **Figure 11** shows the linear regression analysis of the seeding frequency to the evaluation parameters. As shown in **Figure 11**, the relative errors of the qualified index, multiple seeding, miss-seeding index and seeding amount all increased gradually with increasing seeding frequency. At the same time, the relative measurement error was relatively scattered, and the overlap points decreased with increasing seeding frequency. The monitoring stability of the sensor decreased gradually. This may have occurred because under the condition of a fixed seeding speed, the greater the seeding frequency was, the smaller the time interval between the two seeds was. This, in turn, led to the increase in the seed state and the connection between the seeds in the state of the multiple seeding. The probability of multiple seeding and missed seeding increased, and the probability of sensor misjudgment increased.

To clarify the significance of the sensor monitoring accuracy under different seeding frequencies, Design-Expert 8.0.6 software was used to analyze the frequencies of the seeds passing over the sensor, as shown in **Table 4**. The P values of the qualified index relative error, the multiple seeding relative error, the miss-seeding index relative error and the seeding amount relative error were all less than 0.0001. The results showed that the influence of the seeding frequency on the evaluation parameters was very significant.

Linear regression analysis of the seeding speed was carried out. **Figure 12** shows the linear regression analysis of the seeding speed to the evaluation parameters. As shown in **Figure 12**, the relative errors of the qualified index, multiple seeding, miss-seeding index and seeding amount all increased gradually with increasing seeding speed. At the same time, the relative error of measurement was relatively concentrated, and there were more overlap points with increasing seeding speed. Although the monitoring accuracy of the sensor had a downward trend, the stability did not fluctuate greatly with increasing seeding speed. This may be because under the condition of fixed seeding frequency, the greater the seed metering speed was, the faster the opening and closing speed of the seed-pick finger-clip was. This,

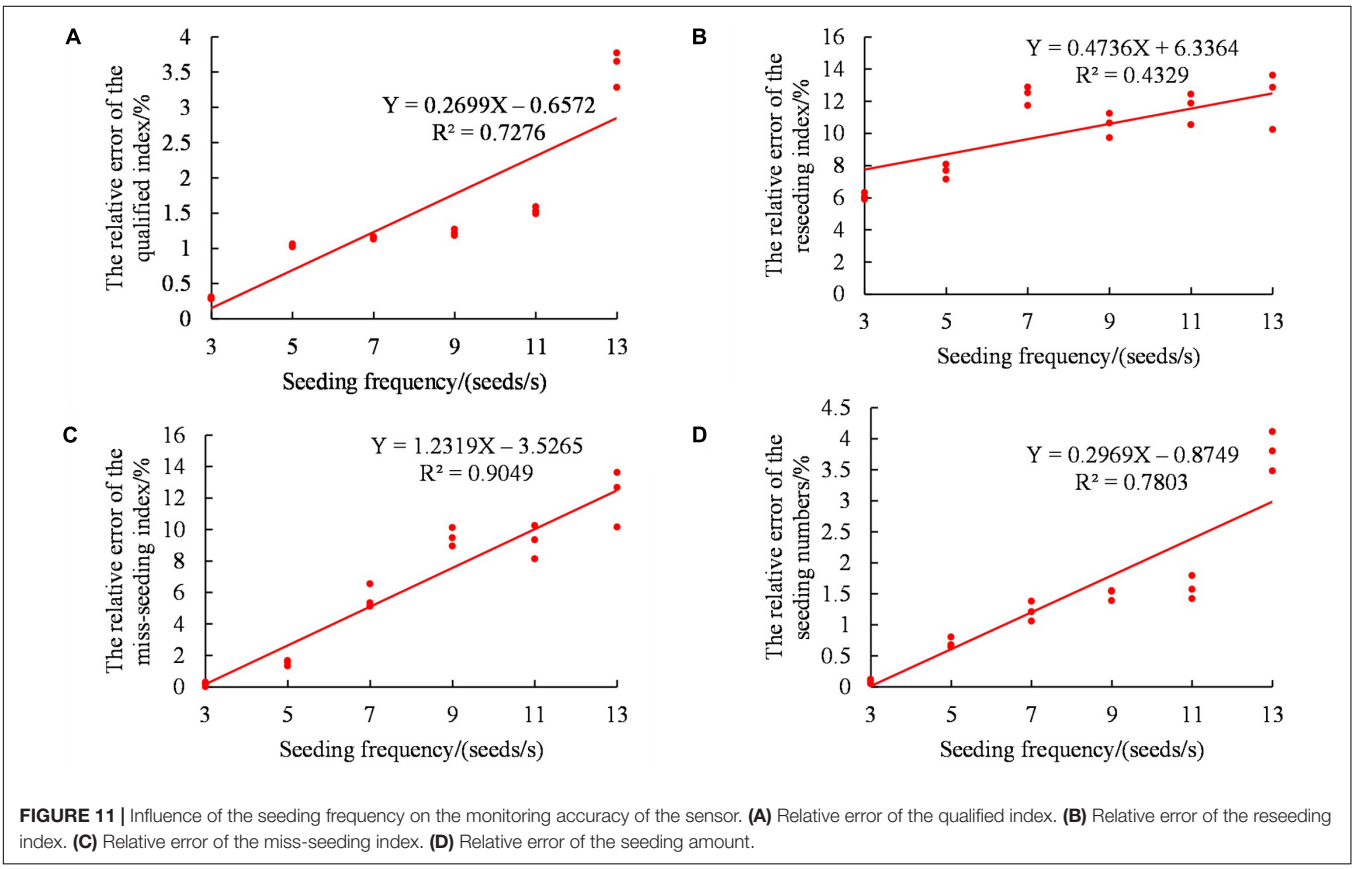


FIGURE 11 | Influence of the seeding frequency on the monitoring accuracy of the sensor. **(A)** Relative error of the qualified index. **(B)** Relative error of the reseeding index. **(C)** Relative error of the miss-seeding index. **(D)** Relative error of the seeding amount.

TABLE 4 | Analysis of seeding frequency variance based on the evaluation parameters.

Evaluation parameters	Source of variation	Sum of squares	df	Mean square	*F*-value	*P*-value
The relative error of the qualified index	Seeding frequency	18.41	5	3.68	312.15	< 0.0001
	Pure Error	0.14	12	0.012		
	Total variation	18.55	17			
The relative error of the reseeding index	Seeding frequency	103.21	5	20.64	23.36	< 0.0001
	Pure Error	10.60	12	0.88		
	Total variation	113.81	17			
The relative error of the miss-seeding index	Seeding frequency	342.12	5	68.42	76.94	< 0.0001
	Pure Error	10.67	12	0.89		
	Total variation	352.79	17			
The relative error of the seeding numbers	Seeding frequency	23.94	5	4.79	163.43	< 0.0001
	Pure Error	0.35	12	0.029		
	Total variation	24.29	17			

in turn, led to a decrease in the stability of the seed from the seed-pick finger-clip to the seed guide belt. The counting accuracy of the diffuse reflection photoelectric sensor and rectangular optical fiber sensor was reduced. This was closely related to the structure and working principle of the finger-clip corn seed metering device.

To clarify the significance of the sensor monitoring accuracies under different seeding speeds, Design-Expert 8.0.6 software was used to analyze the single factor variance in the speed of seeds passing over the sensor, as shown in **Table 5**. The P values of the qualified index relative error, the multiple seeding relative error, the miss-seeding index relative error and the seeding amount relative error were all less than 0.0001. The results showed that the influence of the seeding speed on the evaluation parameters was very significant.

DISCUSSION

As the most widely used mechanical seed metering device, the improvement and effective monitoring of the sowing quality of finger-clip precision corn seed metering device is of great value to promote the development of the sowing link of precision agriculture. The conventional finger-clip precision corn seed metering device is equipped with a seed guide tube under it, so it is unable to sowing at zero speed, resulting in the reduction of seed metering performance when the seeds bounce and collide. In addition, the structural configuration cannot meet the layout requirements of sensors. In this study, the seed guide tube structure was replaced by the lengthened seed guide belt, which simplified the overall structure. At the same time, the research was carried out from the aspect of stable delivery mechanism to achieve the effect of zero speed seed delivery. Due

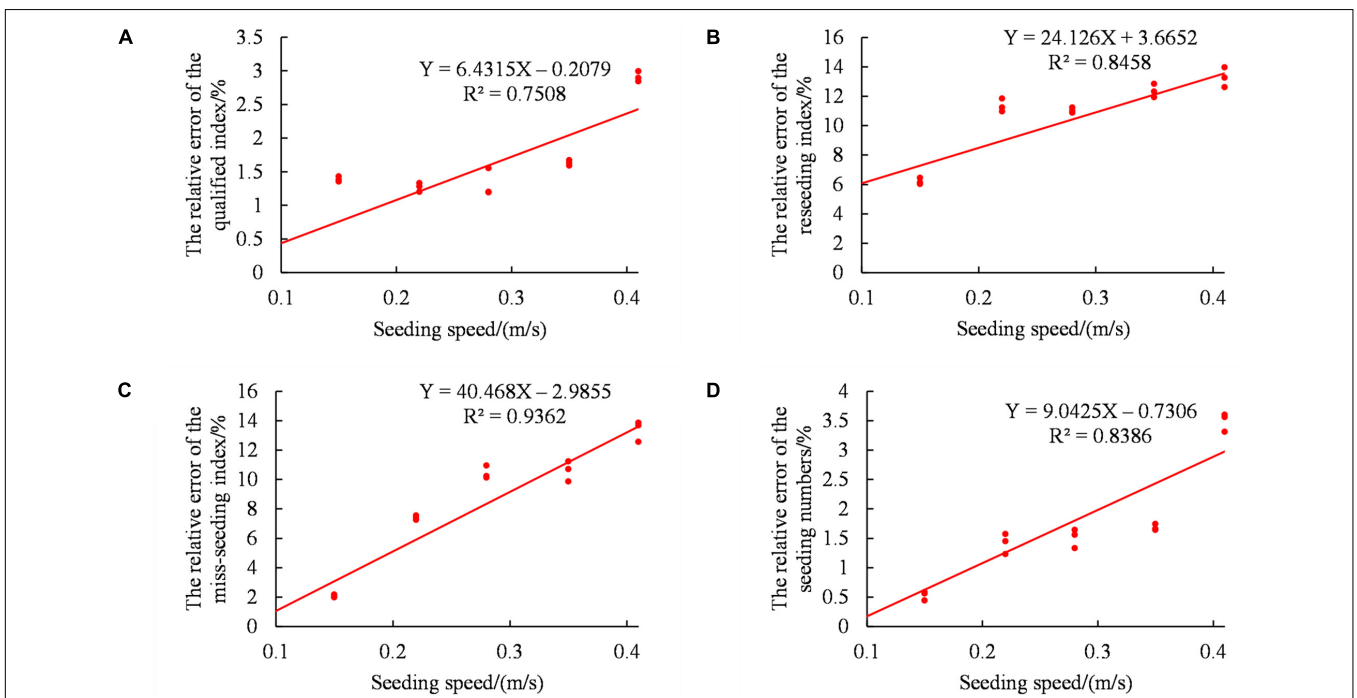


FIGURE 12 | Influence of the seeding speed on the monitoring accuracy of the sensor. **(A)** Relative error of the qualified index. **(B)** Relative error of the reseeding index. **(C)** Relative error of the miss-seeding index. **(D)** Relative error of the seeding amount.

TABLE 5 | Analysis of the seeding speed variance based on the evaluation parameters.

Evaluation parameters	Source of variation	Sum of squares	df	Mean square	*F*-value	*P*-value
The relative error of the qualified index	Seeding speed	12.15	5	2.43	256.43	< 0.0001
	Pure Error	0.11	12	0.009472		
	Total variation	12.26	17			
The relative error of the reseeding index	Seeding speed	170.83	5	34.17	203.29	< 0.0001
	Pure Error	2.02	12	0.17		
	Total variation	172.85	17			
The relative error of the miss-seeding index	Seeding speed	414.15	5	82.83	407.54	< 0.0001
	Pure Error	2.44	12	0.20		
	Total variation	416.59	17			
The relative error of the seeding numbers	Seeding speed	20.79	5	4.16	262.90	< 0.0001
	Pure Error	0.19	12	0.016		
	Total variation	20.98	17			

to the limitation of the structure and metering mode of finger-clip precision corn seed metering device, the traditional single sensor monitoring cannot meet its monitoring requirements. Based on multi-sensor information fusion, a diffuse reflection photoelectric sensor and a rectangular grating sensor were configured at the seed guide and seed feeding positions, respectively, which solved the problem that the traditional single seed metering device cannot judge the seed overlap, and effectively improved the monitoring accuracy. In addition, based on C#, a software interactive interface which was easy for secondary development was designed. Comprehensive bench test and field test verified the effectiveness and accuracy of seed metering quality and monitoring system. This study provides ideas for the optimal design scheme of zero speed seeding at a low position, and provides reference for the development of multi-sensor fusion monitoring system in sowing link.

The performance of mechanical seed metering device is determined by its seed metering mode, seed guiding characteristics and inherent properties. Kocher et al. (2011) pointed out that the uniformity index of seeds is determined by the seed guide tube. The convention finger-clip precision corn seed metering device is equipped with a seed guide tube under it, so it is unable to seed at zero speed, resulting in the reduction of seed metering performance when the seeds bounce and collide. When the rotation speed of the conventional finger-clip precision corn seed metering device was 45r/min, the coefficient of variation was 17.97% (Wang et al., 2017). In this study, the bench test of conventional finger-clip precision corn seed metering device was carried out, and the research results were basically the same, which was 15.51%. The seed guide tube structure was replaced by the lengthened seed guide belt, which simplified the overall structure. At the same time, the research was carried out from the aspect of stable delivery mechanism to achieve the effect of zero speed seed delivery. The coefficient of variation decreased to 5.63% when the speed of seed metering disk was 45 r/min. In addition, the rotation speed of the seed metering disk was increased. When the rotation speed of the seed metering disk was 65 r/min, the qualificated index was 84.7%, the reseeding index was 8.49%, the miss-seeding index was 6.81%, and the coefficient of variation was 6.83%. If the seed metering performance needs to be further improved, the seed metering mode needs to be changed or the reseeding device needs to be added.

If a single sensor is used for monitoring, the partially overlapped adjacent seeds cannot generate photoelectric signal interval pulse, and the system cannot monitor the seed falling time difference, resulting in missed judgment of overlapped seeds. When the rotation speed of seed metering disk was 36 r/min, the monitoring accuracy of missed sowing was 85.6% (Ji et al., 2016). In this study, a seed metering quality monitoring system is developed based on multi-sensor fusion. The high seed metering speed leads to the decline of sensor monitoring accuracy, which is consistent with the change trend of using infrared sensor developed by Karimi et al. (2019). When the rotation speed of seed metering disk was 35r/min, the monitoring accuracy of missed sowing was 92.50%, which effectively improved the monitoring accuracy of missed sowing. In addition, in order to facilitate debugging, the monitoring system collects data through the USB to TTL chip on the upper computer. With the promotion and popularization of precision agriculture, more and more on-board computers will be configured on tractors. Later, this study will be transplanted to on-board computers for practical operation.

The speed of the planter has an important influence on the sowing performance. With the increase of speed, the grain spacing increases, the qualified index decreases and the miss-seeding index increases (Cay et al., 2018; Yang et al., 2019). The speed of planter directly determines the speed of seed tray, which is consistent with the results of this study. The speed of seed tray not only has an impact on seed metering performance, but also has a certain impact on seeds. The higher the speed of the seed tray, the greater the impact on the seed when clamping the seed, which is easy to cause seed damage. Staggenborg et al. (2004) pointed out that the increase of operation speed will also reduce the corn yield. In addition, bench experiment and field experiment was conducted, and the comparison results was shown in **Table 2**. The field experiment has little influence on the accuracy of the monitoring system and has a great influence on the performance of the seed metering device. Under the same rotating speed, the seed metering performance of field experiment was lower than that of bench experiment. This is mainly due to the uneven field ground, complex and uncontrollable vibration, which leads to the change of seed posture in the clamping process, resulting in the phenomenon of reseeding and miss-seeding. In the later stage, we will focus on the in-depth research from the corn seed posture as the starting point to explore the seed damage mechanism under different seed

metering disk speed and the seed falling posture under different vibration frequency. In addition, the growth state of maize from sowing to harvest will be tracked to explore the corn yield under different working conditions.

CONCLUSION

Through the optimization of the conventional finger-clip precision corn seed metering device, a long-belt structure was incorporated in this study. The overall structure and working principle of the long-belt finger-clip precision corn seed metering device were described, the mechanism of stable transportation and delivery was analyzed, and a corn seeding quality monitoring system was designed. The conclusions were as follows:

(1). The seeding performance of the long-belt finger-clip precision corn seed metering device and the conventional finger-clip precision corn seed metering device decreased with increasing speed for the seed metering tray. Compared with the conventional finger clip precision corn metering device, when the rotation speed of the seed metering disk was 65r/min, the qualified index of the long-belt finger-clip precision corn seed metering device increased from 75.75 to 84.70%, the reseeding index decreased from 13.66 to 8.49%, the miss-seeding index decreased from 10.59 to 6.81%, and the coefficient of variation decreased from 20.69 to 6.83%. The long-belt finger-clip precision corn seed metering device can transport and deliver corn seeds smoothly and effectively to reduce the reseeding index, the miss-seeding index and the coefficient of variation.

(2). The relative error of the qualified index, the reseeding index, the miss-seeding index increased with increasing seed metering tray speed. In the field experiment, the relative deviation of qualified index increased from 0.28 to 4.67%, the relative deviation of reseeding index increased from 6.45 to 13.15%, and the relative deviation of miss-seeding index increased from 3.13 to 15.71%. The monitoring accuracy of the sensor decreased gradually.

(3). The effects of the seeding frequency and seeding speed on the four evaluation parameters were very significant. With increasing seeding frequency, the relative errors of the qualified index, the reseeding index, the miss-seeding index and seeding amount all increased gradually, and the monitoring stability of the sensor decreased gradually. With increasing seeding speed, the relative errors of the qualified index, the reseeding index, the miss-seeding index and seeding amount all increased gradually, but the monitoring stability of the sensor was basically unchanged.

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/s.

AUTHOR CONTRIBUTIONS

HT, CX, ZW, and JW designed and performed the experiments and analyzed the data. HT, CX, and QW wrote 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 (NSFC), grant no. 31901414; the Natural Science Foundation of Heilongjiang Province of China for Excellent Youth Scholars, grant no. YQ2021E003; and the Young Talents Project of Northeast Agricultural University, grant no. 19QC41.

ACKNOWLEDGMENTS

We thank the engineering training center of Northeast Agricultural University for providing the test site.

REFERENCES

Al-Mallahi, A. A., and Kataoka, T. (2016). Application of fibre sensor in grain drill to estimate seed flow under field operational conditions. *Comput. Electron. Agric.* 121, 412–419. doi: 10.1016/j.compag.2016.01.006

Besharati, B., Navid, H., Karimi, H., Behfar, H., and Eskandari, I. (2019). Development of an infrared seed-sensing system to estimate flow rates based on physical properties of seeds. *Comput. Electron. Agric.* 162, 874–881. doi: 10.1016/j.compag.2019.05.041

Cay, A., Kocabiyik, H., and May, S. (2018). Development of an electro-mechanic control system for seed-metering unit of single seed corn planters Part II: field performance. *Comput. Electron. Agric.* 145, 11–17. doi: 10.1016/j.compag.2017.12.021

Chen, J., Bian, J., Li, Y., Zhao, Z., and Wang, J. (2009). Performance detection experiment of precision seed metering device based on high-speed camera system. *Nongye Gongcheng Xuebao/Trans. Chinese Soc. Agric. Eng.* 25, 90–95. doi: 10.3969/j.issn.1002-6819.2009.09.016

Ding, Y., Yang, J., Zhu, K., Zhang, L., Zhou, Y., and Liao, Q. (2017). Design and experiment on seed flow sensing device for rapeseed precision metering device. *Nongye Gongcheng Xuebao/Trans. Chinese Soc. Agric. Eng.* 33, 29–36. doi: 10.11975/j.issn.1002-6819.2017.09.004

Haque, M. E., Bell, R. W., Kassam, A., and Mia, M. N. N. (2016). Versatile strip seed drill: a 2-wheel tractor-based option for smallholders to implement conservation agriculture in Asia and Africa. *Environments - MDPI* 3, 1–13. doi: 10.3390/environments3010001

Hossen, M., Musabbir, A., Hossain, M., Hasan, M., and Islam, M. (2014). Improvement of the seed metering device of power tiller operated zero till drill. *Progress. Agric.* 24, 203–210. doi: 10.3329/pa.v24i1-2.19173

Huang, D., Jia, H., Qi, Y., Zhu, L., and Li, H. (2013). Seeding monitor system for planter based on polyvinylidence fluoride piezoelectric film. *Nongye Gongcheng Xuebao/Trans. Chinese Soc. Agric. Eng.* 29, 15–22. doi: 10.3969/j.issn.1002-6819.2013.23.003

Ji, C., Chen, X., Chen, J., Wang, S., and He, P. (2016). Monitoring system for working performance of no-tillage corn precision seeder. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 47, 1–6. doi: 10.6041/j.issn.1000-1298.2016.08.001

Karayel, D., Wiesehoff, M., Özmerzi, A., and Müller, J. (2006). Laboratory measurement of seed drill seed spacing and velocity of fall of seeds using

high-speed camera system. *Comput. Electron. Agric.* 50, 89–96. doi: 10.1016/j.compag.2005.05.005

Karimi, H., Navid, H., Besharati, B., Behfar, H., and Eskandari, I. (2017). A practical approach to comparative design of non-contact sensing techniques for seed flow rate detection. *Comput. Electron. Agric.* 142, 165–172. doi: 10.1016/j.compag.2017.08.027

Karimi, H., Navid, H., Besharati, B., and Eskandari, I. (2019). Assessing an infrared-based seed drill monitoring system under field operating conditions. *Comput. Electron. Agric.* 162, 543–551. doi: 10.1016/j.compag.2019.04.045

Kocher, M. F., Coleman, J. M., Smith, J. A., and Kachman, S. D. (2011). Corn seed spacing uniformity as affected by seed tube condition. *Appl. Eng. Agric.* 27, 177–183.

Kumar, R., and Raheman, H. (2018). Detection of flow of seeds in the seed delivery tube and choking of boot of a seed drill. *Comput. Electron. Agric.* 153, 266–277. doi: 10.1016/j.compag.2018.08.035

Lan, Y., Kocher, M. F., and Smith, J. A. (1999). Opto-electronic sensor system for laboratory measurement of planter seed spacing with small seeds. *J. Agric. Eng. Res.* 72, 119–127. doi: 10.1006/jaer.1998.0353

Li, M., Ding, Y., and Liao, Q. (2010). Loss sowing detection in field of pneumatic precision metering device for rapeseed. *Nongye Gongcheng Xuebao/Trans. Chinese Soc. Agric. Eng.* 26, 27–31. doi: 10.3969/j.issn.1002-6819.2010.z1.007

Liu, Q., Cui, T., Zhang, D., Yang, L., Wang, Y., He, X., et al. (2018). Design and experimental study of seed precise delivery mechanism for high-speed maize planter. *Int. J. Agric. Biol. Eng.* 11, 81–87. doi: 10.25165/ijabe.v11i4.2802

Maleki, M. R., Jafari, J. F., Raufat, M. H., Mouazen, A. M., and de Baerdemaeker, J. (2006). Evaluation of seed distribution uniformity of a multi-flight auger as a grain drill metering device. *Biosyst. Eng.* 94, 535–543. doi: 10.1016/j.biosystemseng.2006.04.003

Miller, E. A., Rascon, J., Koller, A., Porter, W. M., Taylor, R. K., Raun, W. R., et al. (2012). "Evaluation of corn seed vacuum metering systems," in *American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012*, (Dallas, TX), 815–825.

Navid, H., Ebrahimian, S., Gassemzadeh, H., and Mousavi Nia, M. (2011). Laboratory evaluation of seed metering device using image processing method. *Aust. J. Agric. Eng.* 2, 1–4.

Qiu, Z., Zhang, W., Zhao, B., Ji, J., Jin, X., and He, Z. (2019). Design and test of operation quality monitoring system for small grain electric seeder. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 50, 77–83. doi: 10.6041/j.issn.1000-1298.2019.04.009

Soyoye, B. O. (2020). Design and evaluation of a motorized multi-grain crop planter. *Agric. Eng. Int. CIGR J.* 22, 54–66.

Staggenborg, S. A., Taylor, R. K., and Maddux, L. D. (2004). Effect of planter speed and seed firmers on corn stand establishment. *Appl. Eng. Agric.* 20, 573–580.

Wang, G., Sun, W., Zhang, H., Liu, X., Li, H., Yang, X., et al. (2020). Research on a kind of seeding-monitoring and compensating control system for potato planter without additional seed-metering channel. *Comput. Electron. Agric.* 177:105681. doi: 10.1016/j.compag.2020.105681

Wang, J., Tang, H., Wang, J., Shen, H., Feng, X., and Huang, H. (2017). Analysis and experiment of guiding and dropping migratory mechanism on pickup finger precision seed metering device for corn. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 48, 29–37, 46. doi: 10.6041/j.issn.1000-1298.2017.01.005

Wang, J., Zhang, Z., Wang, F., Jiang, Y., and Zhou, W. (2019). Design and experiment of monitoring system for rice hill-direct-seeding based on piezoelectric impact method. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 50, 74–84, 99. doi: 10.6041/j.issn.1000-1298.2019.06.008

Xing, H., Wang, Z., Luo, X., He, S., and Zang, Y. (2020). Mechanism modeling and experimental analysis of seed throwing with rice pneumatic seed metering device with adjustable seeding rate. *Comput. Electron. Agric.* 178:105697. doi: 10.1016/j.compag.2020.105697

Yang, S., Wang, X., Gao, Y., Zhao, X., Dou, H., and Zhao, C. (2019). Design and experiment of motor driving bus control system for corn vacuum seed meter. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 50, 57–67. doi: 10.6041/j.issn.1000-1298.2019.02.007

Yazgi, A., and Degirmencioglu, A. (2014). Measurement of seed spacing uniformity performance of a precision metering unit as function of the number of holes on vacuum plate. *Meas. J. Int. Meas. Confed.* 56, 128–135. doi: 10.1016/j.measurement.2014.06.026

Zhang, J., Chen, H., Ouyang, B., and Ji, W. (2013). Monitoring system for precision seeders based on a photosensitive sensor. *Qinghua Daxue Xuebao/J. Tsinghua Univ.* 53, 265–268, 273.

Zhang, L., Zhao, Z., Ke, X., Feng, X., Zhang, Z., and Huang, J. (2011). Seed-counting system design using piezoelectric sensor. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 42, 41–45.

Zhao, J., Zheng, C., Zhang, J., Han, D., Nian, Y., and Sun, N. (2020). Parameter optimization and experiment of differential filling groove single grain seed-metering device for wheat. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 51, 65–74. doi: 10.6041/j.issn.1000-1298.2020.12.007

Zhao, L., and Jian, X. (2005). Study on optoelectronic sensor for performance detection of a seedmeter. *Nongye Jixie Xuebao/Trans. Chinese Soc. Agric. Mach.* 36, 41–43, 49.

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 02 February 2022
doi: 10.3389/fpls.2021.810113

Hyperspectral Imaging With Machine Learning to Differentiate Cultivars, Growth Stages, Flowers, and Leaves of Industrial Hemp (*Cannabis sativa* L.)

Yuzhen Lu[1], Sierra Young[2]*†, Eric Linder[3], Brian Whipker[4] and David Suchoff[3]**

[1] Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, MS, United States, [2] Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, NC, United States, [3] Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC, United States, [4] Department of Horticultural Science, North Carolina State University, Raleigh, NC, United States

As an emerging cash crop, industrial hemp (*Cannabis sativa* L.) grown for cannabidiol (CBD) has spurred a surge of interest in the United States. Cultivar selection and harvest timing are important to produce CBD hemp profitably and avoid economic loss resulting from the tetrahydrocannabinol (THC) concentration in the crop exceeding regulatory limits. Hence there is a need for differentiating CBD hemp cultivars and growth stages to aid in cultivar and genotype selection and optimization of harvest timing. Current methods that rely on visual assessment of plant phenotypes and chemical procedures are limited because of its subjective and destructive nature. In this study, hyperspectral imaging was proposed as a novel, objective, and non-destructive method for differentiating hemp cultivars, growth stages as well as plant organs (leaves and flowers). Five cultivars of CBD hemp were grown greenhouse conditions and leaves and flowers were sampled at five growth stages 2–10 weeks in 2-week intervals after flower initiation and scanned by a benchtop hyperspectral imaging system in the spectral range of 400–1000 nm. The acquired images were subjected to image processing procedures to extract the spectra of hemp samples. The spectral profiles and scatter plots of principal component analysis of the spectral data revealed a certain degree of separation between hemp cultivars, growth stages, and plant organs. Machine learning based on regularized linear discriminant analysis achieved the accuracy of up to 99.6% in differentiating the five hemp cultivars. Plant organ and growth stage need to be factored into model development for hemp cultivar classification. The classification models achieved 100% accuracy in differentiating the five growth stages and two plant organs. This study demonstrates the effectiveness of hyperspectral imaging for differentiating cultivars, growth stages and plant organs of CBD hemp, which is a potentially useful tool for growers and breeders of CBD hemp.

Keywords: industrial hemp, classification, hyperspectral imaging, image processing, machine learning

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Yang Li,
Shihezi University, China
Roozbeh Rajabi,
Qom University of Technology, Iran

****Correspondence:***
Yuzhen Lu
yzlu@abe.msstate.edu
Sierra Young
sierra.young@usu.edu
David Suchoff
dhsuchof@ncsu.edu

†Present address:
Sierra Young,
Department of Civil and Environmental Engineering, Utah State University, Logan, UT, United States

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *06 November 2021*
Accepted: *27 December 2021*
Published: *02 February 2022*

Citation:
Lu Y, Young S, Linder E, Whipker B and Suchoff D (2022) Hyperspectral Imaging With Machine Learning to Differentiate Cultivars, Growth Stages, Flowers, and Leaves of Industrial Hemp (Cannabis sativa L.). Front. Plant Sci. 12:810113. doi: 10.3389/fpls.2021.810113

INTRODUCTION

Industrial hemp, or briefly known as hemp, is a crop cultivated for producing a wide range of industrial and consumer products (Renée, 2018). Hemp belongs to the same plant species (*Cannabis sativa* L.) as marijuana that is mainly used recreationally for its intoxicating properties. In the United States, hemp is legally defined as *Cannabis sativa* L. that contains no more than 0.3% total tetrahydrocannabinol (THC), the compound that is responsible for getting a person high and more abundant in marijuana. Because of its association with marijuana, commercial production of hemp in the United States has been long restricted until the passage of the 2018 Farm Bill (Schluttenhofer and Yuan, 2019). As of 2021, all the states in the United States have legalized hemp production for commercial or research purposes. There are three main types of hemp that are grown for different markets, i.e., fiber, oilseed, and cannabidiol (CBD) (Cherney and Small, 2016; Adesina et al., 2020), among which CBD demand is currently the driving force for hemp growth (Carpenter and Peroutek, 2019). While the medicinal uses of CBD are still being researched, market opportunities for CBD hemp are expected to be significant, with CBD sales in the United States projected to reach $23.7 billion by 2023 (Brightfield Group, 2019).

Due to the potential of CBD hemp as an economically viable crop, many farmers are turning to hemp as an alternative crop to fit into their current production system and utilize established farm infrastructure. In a recent survey conducted among North Carolina organic farmers, 85% of the growers expressed interest in growing hemp on their farms and the vast majority intended to grow hemp primarily for CBD (Dingha et al., 2019). As an emerging cash crop, many uncertainties surround producing hemp profitably (Adesina et al., 2020), such as cultivar selection, transplanting dates, planting densities, fertilization, pest management, and harvest dates. Confounding these uncertainties is the federal regulatory limit of THC. Production of hemp with THC levels above 0.3% in the United States can mean the destruction of hundreds of acres and loss of thousands of dollars (USDA-AMS, 2021), which could have been avoided through proper cultivar/variety selection and improved production practices. Hence there is a practical need to identify and discriminate hemp phenotypes and cultivars to facilitate crop management as well as serving forensic purposes. The growth stage of hemp at harvest time, in addition to genetics and environmental factors of seed stocks (Campbell et al., 2019; Glivar et al., 2020), is also an important factor influencing chemical profiles (e.g., THC and CBD) of the plant (De Backer et al., 2012; Stack et al., 2021). It is thus also important to determine growth stages and establish harvest timing recommendations to maximize CBD contents in hemp.

Hemp cultivars and growth stages can be determined by agronomic experts who visually inspect morphological characteristics (e.g., shape, color, and texture) of the plant organs (e.g., leaves and flowers). Visual inspection is affected by inconsistency and variability associated with the perception of inspectors, which is further complicated by significant biological variations within and among hemp cultivars. Some hemp cultivars may not be visually distinct and readily differentiated from each other. Thus, analytical methods, such as gas/liquid chromatography and mass spectrometry (Capriotti et al., 2021), have been proposed for differentiating hemp cultivars based on the chemical fingerprints of the plants (Jin et al., 2017; Wang et al., 2018; Dong et al., 2019). Although accurate and reliable, these methods are slow, costly, require sample preparation and destructive wet-chemistry procedures, and thus are not suitable for rapid, on-site testing applications. Therefore, it would be beneficial if a rapid, non-destructive, and objective method is developed for the differentiation of hemp cultivars as well as growth stages.

Optical sensing technology, which interrogates biological materials non-destructively, is considered an attractive means for addressing the shortcomings of human inspection and analytical testing. Numerous studies have been conducted on using spectroscopic techniques for cultivar/variety differentiation of plants and agricultural products (Cozzolino et al., 2003; Luo et al., 2011; Lu et al., 2014). Recently, Sanchez et al. (2020) used Raman spectroscopy for differentiating hemp, cannabis, and CBD-rich hemp with 100% accuracy. Raman measurements, however, require direct contact of samples with the spectrometer to obtain high-quality signals (Sanchez et al., 2020). Duchateau et al. (2020) used near-infrared spectroscopy for discriminating legal and illegal hemp, defined by a cut-off concentration of 0.2% THC in European Union countries, obtaining classification accuracies of 91–95%. Crushing dried hemp plants was required prior to the spectroscopic measurements (Duchateau et al., 2020). Cirrincione et al. (2021) reported on using attenuated total reflectance infrared spectroscopy for the discrimination between fiber-type and drug-type cannabis samples. Spectroscopic sensing, however, only measures small portions of plant tissues and often requires sample treatments (e.g., drying and grinding) (Duchateau et al., 2020) and direct contact between samples and the detector (Sanchez et al., 2020).

Hyperspectral imaging is a power modality for measuring spectral and spatial information of samples simultaneously (Lu et al., 2020). Compared to spectroscopic techniques that are used for point measurements, hyperspectral imaging is advantageous in delivering reliable and comprehensive analysis of characteristics or properties of plant materials with minimal sample preparation, requiring no sample contact, and thus is potentially more suitable for high-throughput, on-site testing. Pereira et al. (2020) investigated hyperspectral imaging for identifying hemp leaves under natural conditions, achieving sensitivity, and specificity values of 89.45% and 97.60%, respectively. A similar study was conducted by Holmes et al. (2020) on classifying flowers, stems and leaves of hemp using hyperspectral imaging. So far, to the best of our knowledge, no research has been carried out on using hyperspectral imaging for classifying for cultivars and growth stages of CBD hemp.

Given the limitations of existing methods using visual assessment and chemical analysis for phenotyping and characterization of hemp plant materials, the objective of this research is therefore to present a proof-of-concept validation of a novel hyperspectral imaging-based approach for non-destructive, fast, and objective differentiation of cultivars, growth stages and plant organs (i.e., leaves and flowers) of CBD hemp. Specifically,

in this research we acquired hyperspectral images from freshly harvested leaf and flower materials of five cultivars of CBD hemp at five growth stages using a benchtop hyperspectral reflectance imaging system, developed an image processing pipeline for segmenting the plant parts from background and extracting spectra from sample segments, performed exploratory analysis of spectral features of hemp samples, and built classification models to differentiate the cultivars, growth stages, and plant parts. This study demonstrates the efficacy of hyperspectral imaging technology as a tool to differentiate cultivars, growth stages and plant parts of CBD hemp, which will be beneficial for hemp cultivation and breeding programs.

MATERIALS AND METHODS

Hemp Samples

Five CBD hemp cultivars were used in this study, including Cherry Wine (CW), BaOx (BX), First Light 58 (FL58), First Light 70 (FL70), and TJ's (TJ). These cultivars were chosen as they were used in a complimentary field trial to determine the optimum harvest date, and particularly BX and CW represent the majority of CBD hemp cultivars planted in North Carolina. The hemp plants were grown in a greenhouse, as shown in **Figure 1** (left), at the NC State University Horticulture Field Laboratory (Raleigh, NC, United States). The trial was arranged in a complete randomized design containing four replicates. A total of 20 plants (5 harvest dates × 4 replicates) per cultivar were randomly placed on greenhouse benches (total 100 plants for five cultivars).

Hemp harvests took place during September to November of 2020, at 2, 4, 6, 8, and 10 weeks after flower initiation, corresponding to five plant growth stages. At the time of harvest or growth stage, four plants were randomly chosen per cultivar row, corresponding to four replications, and both leaves and flowers, as shown in **Figure 1** (right), were sampled for the differentiation of hemp cultivars and growing stages. For each plant, 4 leaves were sampled from its main apical meristem, and 4–6 flowers were sampled depending on the size and number of flowers on the plant. The details of sample numbers are summarized in **Table 1**. The freshly harvested samples were immediately scanned by a hyperspectral imaging system as described below.

Hyperspectral Image Acquisition

A portable, benchtop hyperspectral reflectance imaging system (**Figure 2**) under controlled lighting was assembled for acquiring images from hemp samples. The system mainly consisted of a line-scan hyperspectral camera (Pika XC2, Resonon Inc., Bozeman, MT, United States), attached with a focusing lens (Xenoplan 1.4/17, Schneider Kreuznach, Bad Kreuznach, Germany), a four-fixture, 140-W halogen lamp assembly (symmetrically oriented with respect to the camera) for providing illumination over samples, a motorized stage (Resonon Inc., Bozeman, MT, United States) and a Spectralon reference target (SRT-20-020, Labsphere, Inc., North Sutton, NH, United States) with nominal reflectance of 20%. Synchronized with the camera, the stage moved a flat sample-holding tray (at a speed of 1 cm/s) for hyperspectral line scanning. The reference, which was placed on the tray and scanned along with samples, as shown in **Figure 2** (right), was used for standardizing the spectral responses of the camera. The imaging system was operated in an enclosed chamber to prevent interference from ambient light.

Image acquisitions were conducted on five harvest occasions as indicated above. The hemp leaves and flowers were imaged separately for individual plants. The software SpectrononPro (Resonon Inc., Bozeman, MT, United States) was used for controlling the camera and motorized stage during imaging. The acquired hyperspectral datacube consisted of 462 wavelengths over a wavelength range of 400–1000 nm (at a spectra resolution of 1.3 nm), and spatially each scanning line consisted of 1,600 pixels (at a spatial resolution about 0.5 mm^2 per pixel for hemp samples), and the number of scanning lines per datacube depended on the actual scanning duration.

Image Processing

The acquired hyperspectral images were processed to segment the reference and hemp samples from the background. Thresholding is a simple and effective technique for image segmentation, provided that the image histogram has well-defined modes corresponding to regions of interests. While a flat, uniformly colored tray (**Figure 3**) was used as the background for hemp imaging, there was still noticeable illumination unevenness in acquired images (**Figure 3**), restricting using a global threshold for object segmentation. To facilitate the segmentation, a robust algorithm was developed by obtaining a contrast-optimized, normalized band difference (NBD) image, followed by applying an INTERMODE thresholding technique (Glasbey, 1993; Lu and Lu, 2017). NBD is calculated as vegetation indices in hyperspectral sensing to improve feature discrimination; it can be defined as in a general form (Ferwerda et al., 2005). $\boldsymbol{I = (R_{@\lambda 1} - R_{@\lambda 2}) / (R_{@\lambda 1} + R_{@\lambda 2})}$, where, $\boldsymbol{R_{@\lambda 1}}$ and $\boldsymbol{R_{@\lambda\ 1}}$ denote the reflectance images at wavelengths λ1 and λ2 (λ1 > λ2), respectively.

In this study, the best wavelength pair was determined, as illustrated in **Figure 3**, by calculating NBD images for all waveband pairs and choosing the one at which the maximum image contrast is obtained (Lu et al., 2021). The image contrast was defined as the ratio of among-class (plant pixels vs. non-plant pixels) variance to the total variance of an image, following the principle of the Otsu's thresholding (Otsu, 1979). As such two wavelengths 898 nm and 474 nm, in near-infrared and blue regions, respectively, were identified for calculating NBD images. It is noted that the algorithm was applied to a single hyperspectral image and the identified wavelength pair was then generalized to all other images. As shown in **Figure 3**, the NDB image is highly contrasted between hemp samples and the background.

The contrast-optimized NBD images enable the segmentation of hemp samples and reference by global thresholding. The histogram of the NBD image, as showed in **Figure 4** (left), has two distinct peaks, and the one at the lower end of the histogram corresponds to the background and the other corresponds to hemp samples and the edge of the reference. The INTERMODE thresholding technique finds the optimal threshold by taking the average of the two peaks of a bimodal histogram (Glasbey, 1993;

FIGURE 1 | Hemp plants grown in a greenhouse **(left)**, and flower and leaf for sampling **(right)**.

TABLE 1 | Sample numbers for five hemp cultivars at different growth stages (sampling dates).

Sampling date	Cherry Wine		BaOx		First light 58		First light 70		TJ's	
	Leaf	Flower	Leaf	Flower	Leaf	Flower	Leaf	Flower	Leaf	Flower
09/24/2020	16	24	16	22	16	16	16	16	16	16
10/08/2020	16	24	16	21	16	17	16	16	16	16
10/22/2020	16	24	16	21	16	16	16	16	16	16
11/05/2020	16	24	16	22	16	16	16	16	16	16
11/19/2020	16	24	16	21	16	16	16	16	16	16

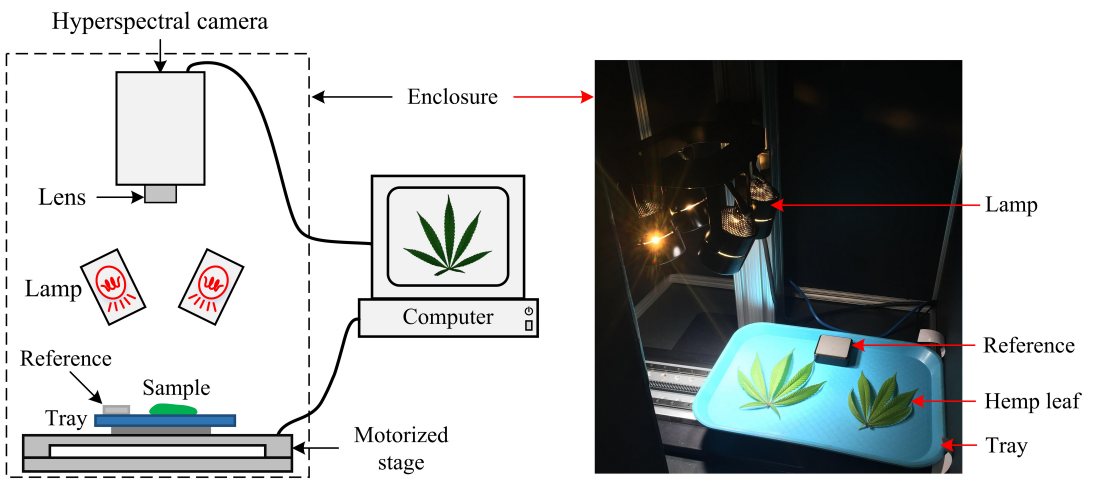


FIGURE 2 | Schematic **(left)** and photograph **(right)** of a hyperspectral imaging system for acquiring images from hemp samples.

Lu and Lu, 2017). Since the raw histogram might not be ideally bimodal, it was subjected to average smoothing iteratively using a three-point window, before determining the optimal threshold, until the smoothed histogram became bimodal. The thresholding was then followed by routine morphological operations to refine the initial segmentation. **Figure 4** (right) shows an example of the segmented hemp leaves and reference.

Furthermore, given a hyperspectral datacube for each scan, mean spectra were extracted for the reference and individual hemp leaves/flower samples, respectively, by averaging the spectra of all the pixels in the corresponding region of interest. Thereafter, ratio spectra were obtained by dividing the spectra of hemp samples by the spectrum of the reference in the same scan, to standardize the spectral responses of the camera, and used for building discriminative models as described below. While morphological or texture features can also be extracted and fused with the mean spectra for modeling tasks, only the latter were used for the modeling tasks and found adequate for yielding high classification accuracies.

Model Development

Machine learning models were developed to differentiate five hemp cultivars, five growth stages (corresponding to five sampling dates) and two plant organs (i.e., leaves and flowers), respectively. For cultivar differentiation, the models were built

Algorithm: find a contrast-optimized normalized band difference image $\mathbf{I}_{opt}$

Input: a hyperspectral matrix $\mathbf{X}$ of size $m \times n \times k$, where the third dimension is wavelength
Initialization: contrast matrix $\mathbf{C} = \mathbf{0}$
  for i=1; $i \le m$
    for j=1; $j \le i$ & $j \le n$
      $\mathbf{I} = (\mathbf{X}(:, :, i) - (\mathbf{X}:, :, j))/(\mathbf{X}(:, :, i) + (\mathbf{X}: , :, j))$
      $\mathbf{C}(i, j) =$ **computImageContrast(I)**
    end
  end
  $r, c =$ **arg max**$(\mathbf{C}(i, j))$
$\mathbf{I}_{opt} = (\mathbf{X}(:, :, r) - (\mathbf{X}:, :, c))/(\mathbf{X}(:, :, r) + (\mathbf{X}: , :, c))$

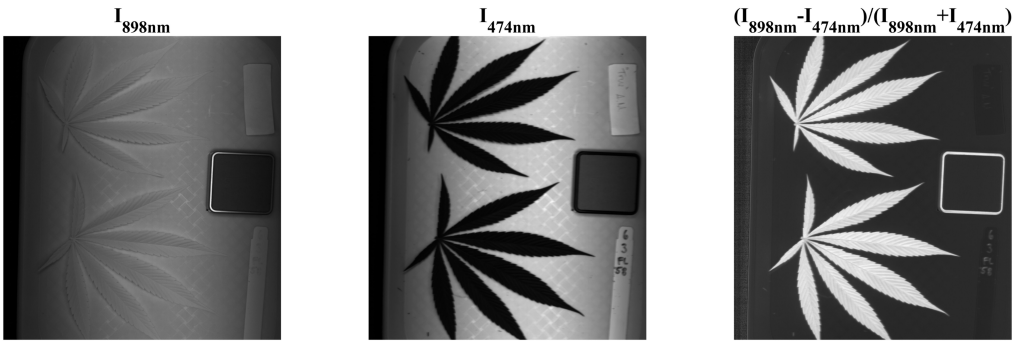


FIGURE 3 | (Top) Algorithmic procedures of finding a contrast-optimized, normalized band difference image from a hyperspectral datacube. **(Bottom)** The images at 898 nm and 474 nm and the corresponding normalized band difference image that exhibits the optimal contrast.

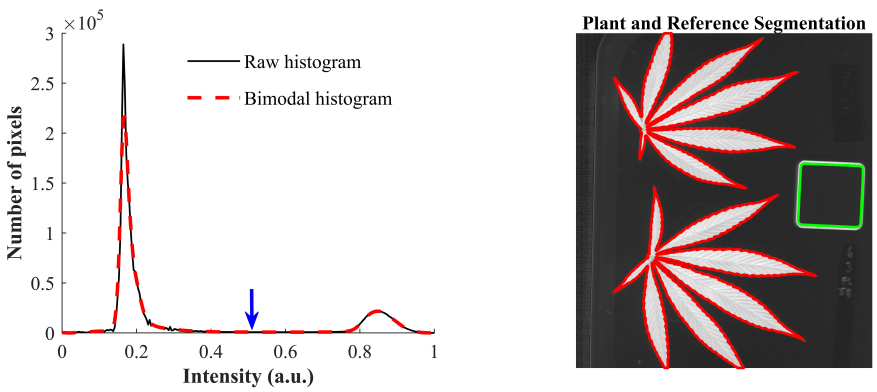


FIGURE 4 | (Left) Raw histogram and bimodal histogram of a contrast-optimized normalized band difference image (**Figure 3**). The blue downward arrow indicates the optimal threshold that corresponds to the mean position of the two peaks of the bimodal histogram that is obtained by iteratively smoothing the raw histogram until it is bimodal. **(Right)** Segmentation of hemp leaves and a reference using the optimal threshold.

using the spectra of hemp leaves and flowers, respectively, as well as using the ensemble of hemp leaf and flower samples, at each growth stage. Furthermore, the ensemble of the samples from different stages is also examined for model development. For growth stage differentiation, similarly, classification models were built using hemp leaves and flowers, respectively, for each cultivar. Moreover, models were built for discriminating hemp leaves and flowers for each cultivar at each growth stage. In each modeling scenario, the spectral dataset was randomly partitioned into training and test sets according to a ratio of 3 to 1 (**Figure 5**), for model training and testing, respectively, and wavelength-wise data normalization was performed so that the reflectance values at each wavelength had a zero mean and a unit variance.

Regularized linear discriminant (rLDA) proposed by Guo et al. (2007) is an extension to the classic LDA specifically for solving classification problems with high-dimensional data. By regularizing the covariance matrix and thresholding (shrinking) the linear coefficients, rLDA is sufficiently robust for modeling high-dimensional data and also very competitive to other far more computation-expensive classifiers such as support vector machine (Guo et al., 2007). Hence, rLDA was chosen for the modeling tasks in this study. There are two tunable hyperparameters in rLDA, i.e., the regularization parameter γ ($0 < \gamma < 1$) and threshold (or shrinkage) parameter δ ($\delta > 0$). When there are more predictors (variables) than samples, which typically holds true for high-dimensional data, the optimal

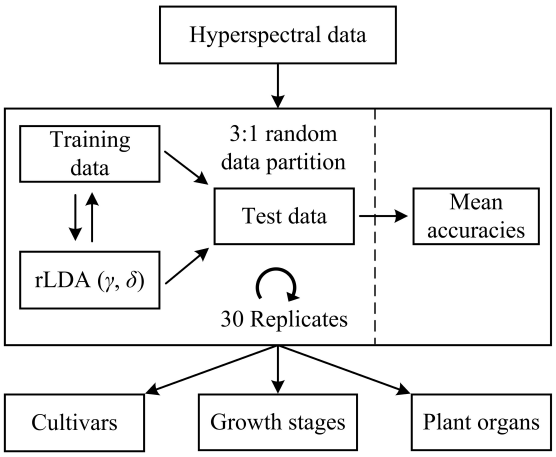


FIGURE 5 | Discriminative modeling of hyperspectral data using regularized linear discriminant analysis (rLDA) for classifying hemp cultivars, growth stages, and plant organs (leaves vs. flowers). For rLDA, two hyperparameters γ and δ are optimized through Bayesian optimization based on 10-fold cross validations during model training.

value of γ is shifted toward 0 (Guo et al., 2007); a higher value of δ implies fewer variables incorporated into the model, which has the effect of variable selection for modeling. Here the two hyperparameters were determined through Bayesian optimization (Snoek et al., 2012) in the context of 10-fold cross validations on the training data, over a range of [0, 0.01] and [1e-3, 1e3] for γ and δ, respectively. The search ranges were chosen based on preliminary testing. Because of the randomness of the spectral data partition, it would be desirable to repeat the modeling procedures multiple times with random dataset partition for obtaining a reliable estimate of model performance. In this study, a repeated holdout validation strategy (also referred to as Monte Carlo cross validation) was performed to avoid potential pitfalls of single data partition (Raschka, 2018). Given the efficiency of rLDA, a relative high number of 30 modeling replications were conducted (**Figure 5**), and the resultant mean value of the classification accuracies (the percentages of the number of correctly classified samples of the total sample number) on the test data was computed for model evaluation. Further, statistical comparisons were conducted on the mean classification accuracies among different models using Fisher's least significant procedure at the 5% significance level.

All the analyses for image preprocessing, feature extraction and model development were performed in Matlab R2020b (The Mathworks, Inc., Natick, MA, United States).

RESULTS AND DISCUSSION

Exploratory Analysis

Figure 6 shows the spectra of hemp samples of the five different cultivars harvested 4 weeks after flower initiation and the spectra of one cultivar at all the growth stages (2–10 weeks after flower initiation). Like other green plants, the spectra of hemp leaves and flowers are characterized by low reflectance in the visible range due to absorption of plant pigments, and reflectance rising rapidly at wavelengths around 700 nm and plateauing in the NIR region, due to reduced absorption and increased scattering of plant tissues in the region (Horler et al., 1983). The major reflectance valley (i.e., the absorption peak) occurring around 670 nm is attributed to the absorption of chlorophylls. Large spectra variations are observed in the visible (450–650 nm around the green band) and NIR regions, among the samples harvested at different plant growth stages. These variations are associated with the dynamics in the chemical profiles (e.g., pigments and water) of plant organs as the plant matures. It seems more apparent that spectral reflectance of hemp leaves increased with the plant age, which is probably due to leaf senescence-induced degradation of chlorophylls (Merzlyak et al., 1999). Water loss accompanying senescence of plant tissues also contributes to increased reflectance in the NIR range (Hunt and Rock, 1989), which may explain high NIR reflectance of hemp flowers in the last scan.

The hemp samples cannot be directly distinguished for the five cultivars from the spectral profiles, because of strong overlapping (**Figure 6** top); whereas there are more noticeable differences in the spectral profiles among the five growth stages (**Figure 6** bottom), and between hemp leaves and flowers. To visualize the distribution of hemp samples of different cultivars, principal component analysis (PCA) was performed on the spectral data. **Figure 7** shows an example of the scatter plots for the hemp samples shown in **Figure 6**. The first two principal components (PCs) account for 87.0% and 91.3% of the total variance of the spectral data of leaf and flower samples, respectively. The scatter plots allow visualizing the unsupervised separation of samples of different classes. Clearly, the samples of five cultivars do not form distinct, well-separated clusters in the PC space, which is also true in the plots by the top three PCs that explain over 96% of the total variance (3D scatter plots are not presented). Similar findings are also observed for the samples harvested at other growth stages. In contrast, the hemp samples among the different growth stages, and especially between plant organs (leaves and flowers) form better-resolved separated groups in the PC space as shown in **Figure 8**. Despite the qualitative analysis, this result suggests that hemp cultivars would not be readily discriminated though unsupervised analysis, highlighting the need for potent supervised classification techniques to distinguish hemp cultivars, and that high accuracies would be achieved in classifying the growth stages and plant organs.

Classification

The machine learning models based on rLDA were first developed for differentiating the hemp cultivars using the spectral data of leaf and flower samples separately as well as their combination, at each growth stage (harvest dates). In each scenario, discriminative models were built and tested over 30 replications with random dataset partition for each replication, and the mean overall classification accuracy on test data was calculated and used as the metric of classification performance.

Figure 9 shows the classification accuracy, with statistical comparisons made between the accuracies at each growth stage.

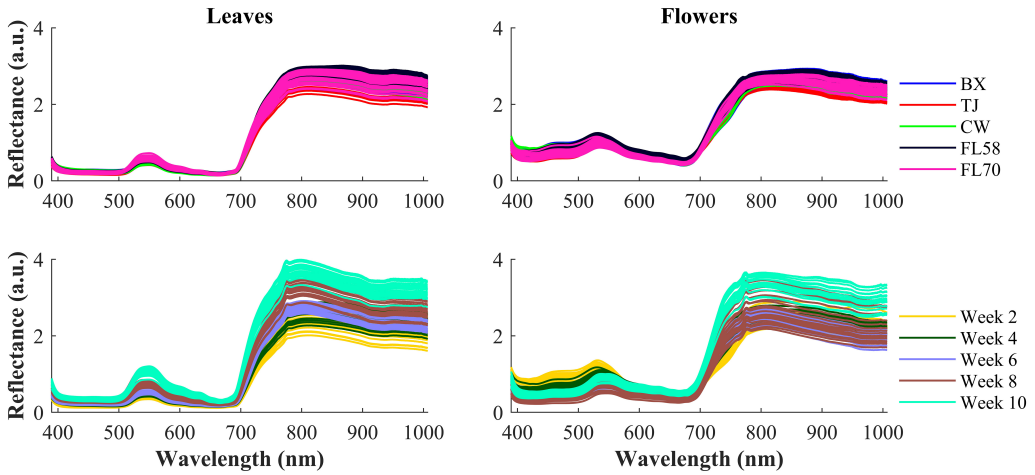


FIGURE 6 | (Top) Mean spectra of five cultivars (i.e., BX, TJ, CW, FL58, and FL70) of hemp flowers and leaves harvested 4 weeks after flower initiation and **(Bottom)** mean spectra of the cultivar Cherry Wine harvested at all the five growth stages. The reflectance (a.u.) is a relative quantity obtained by diving a sample spectrum by that of the standard reference (section "Hyperspectral Image Acquisition") with nominal reflectance of 20%.

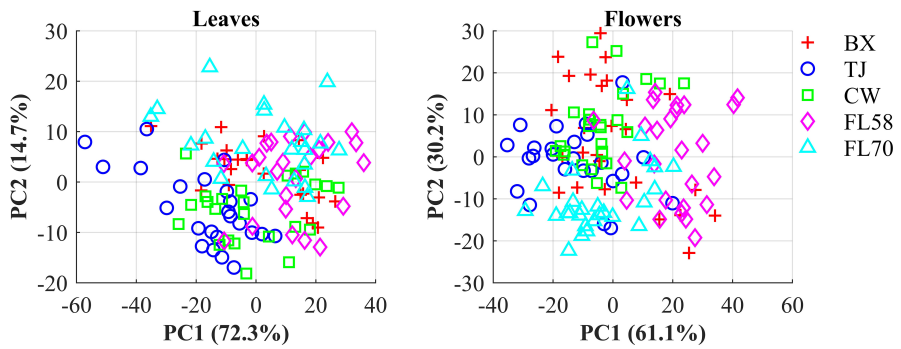


FIGURE 7 | Scatter plots in the space spanned by the first two principal components (PCs) for five cultivars of hemp flowers and leaves 4 weeks after flower initiation (**Figure 6**). The percentage value in parentheses indicates the variance portion explained by the corresponding PC.

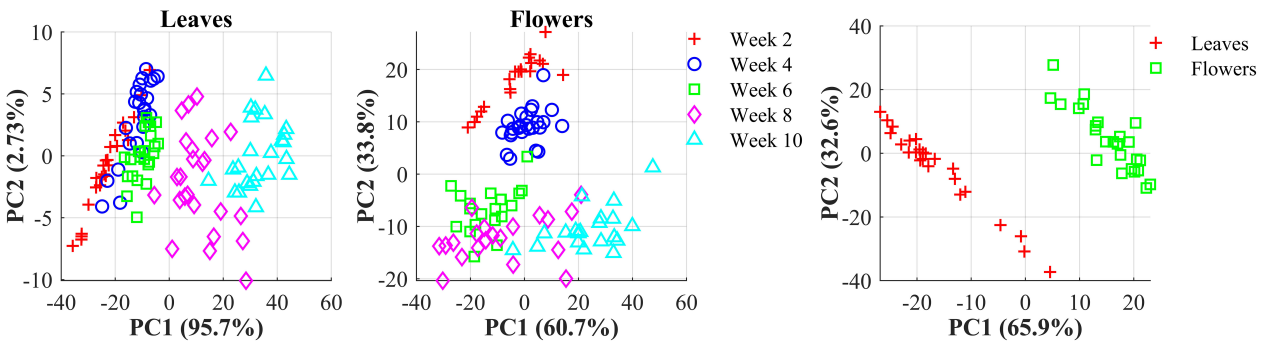


FIGURE 8 | (Left) Scatter plots in the space spanned by the first two principal components (PCs) for hemp sample at five growth stages (2–10 weeks after flower initiation). **(Right)** Scatter plots of the first two PCs for hemp flowers and leaves harvested week 4 after flower initiation. The percentage value in parentheses indicates the variance portion explained by the corresponding PC.

Although the PCA of leaf and flower spectra could not reveal a good separation among different hemp cultivars, the rLDA models based on the leaf or flower samples achieved high classification accuracies ranging 96.8% to 99.6%, with standard errors of less than 1%. The classification accuracies vary with plant organ and growth stage. The leaf samples yielded a

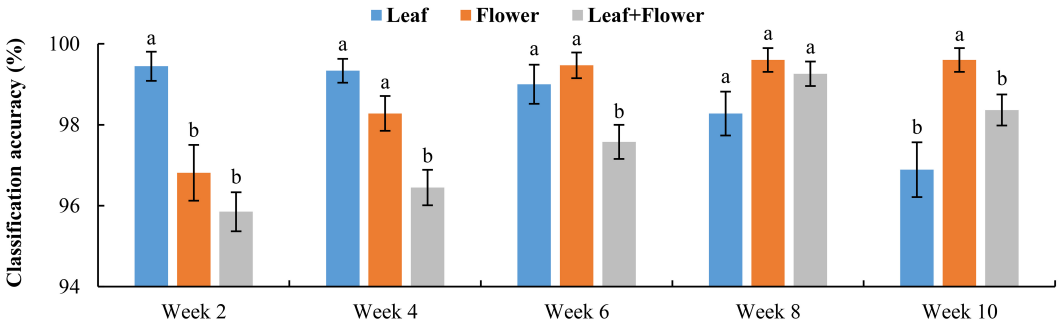


FIGURE 9 | Classification accuracies in differentiating hemp cultivars based on the samples at each growth stage. The classification accuracy is obtained by averaging the accuracies in 30 modeling replicates, and the error bar indicates the corresponding to the (positive/negative) standard error. At each growth stage, the two accuracies with different letters are statistically different at the 5% significance level.

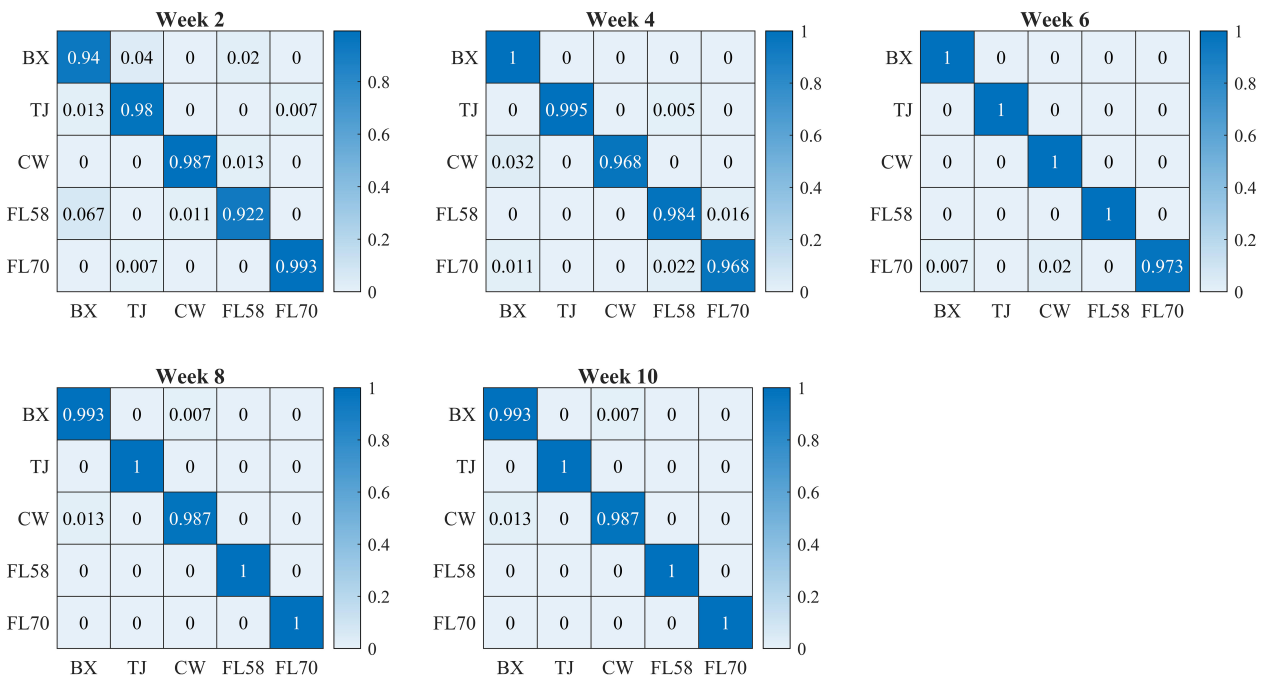


FIGURE 10 | Confusion matrices (rows and columns correspond to true and predicted labels, respectively) for hemp cultivar classification based on the hyperspectral data of flower samples. Each confusion matrix is obtained by pooling and row-wise normalizing classification results over 30 modeling replications.

significantly (+2.6%) better accuracy than that obtained by the flowers at the first growth stage, but a significantly (−2.7%) lower accuracy at the last stage. At the three intermediate stages, the accuracies by the leaves and flowers were similar. It is interesting to note that the accuracy obtained by the leaf samples exhibited a decreasing trend with growth stage, as opposed to an increasing trend for the accuracy by the flowers. The reason underlying this phenomenon has not been fully understood. At week 2, the earliest harvest stage (2 weeks after flower initiation), the flower buds were tiny (3–5 mm) and sticky, which could cause sampling errors. The chemical components (e.g., cannabinoids and cellulose) that are found to be indicative of hemp cultivars (**Sanchez et al., 2020**) may have low concentrations at this stage, which remains to be validated by a further study on chemical analysis of hemp samples. At later growth stages (8–10 weeks after flower initiation), a few hemp plants had minor spider mite (*Tetranychus urticae*) infestation on the leaves, which could also confound the discrimination of hemp cultivars.

Figure 10 shows confusion matrices for cultivar classification based on the hyperspectral data of flower samples. Each of the matrices is obtained by pooling and row-wise normalizing the classification results on test data for 30 modeling replications. The overall classification accuracies are similar among hemp cultivars, and there is no consistent pattern of the most or least correctly classified cultivars over the five growth stages. Similar findings were also observed for models based on the data of hemp leaves and the ensemble of flower and leaf samples (confusion matrices not presented). Compared to the results attained using

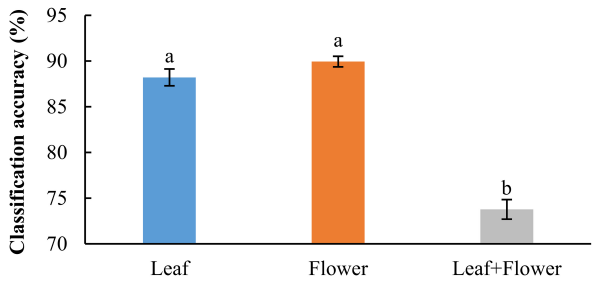


FIGURE 11 | Classification accuracies in differentiating hemp cultivars by pooling samples at all the growth stages. The classification accuracy is obtained by averaging the accuracies in 30 modeling replicates, and the error bar indicates the corresponding to the (positive/negative) standard error. The two accuracies with different letters are statistically different at the 5% significance level.

the leaf and flower samples separately, the combined data of yielded statistically diminished or similar accuracy with the lowest and highest values of 95.8% and 99.3% at week 2 and 8, respectively (**Figure 9**).

In addition to modeling the samples at individual growth stages, the samples collected from different growth stages were also pooled together to build models for cultivar classification. Here, three types of models were built by pooling all the leaf samples, the flower samples and their combination at the five growth stages, resulting in the accuracies of 91.9%, 91.8%, and 82.8%, respectively, as shown in **Figure 11**. The combination of leaf and flower samples led to a significantly lower accuracy than modeling them separately. Compared to the results of models for individual growth stages (**Figure 9**), the accuracies obtained from pooling the samples across growth stages resulted in a marked accuracy reduction of 5.03% to 16.4%. This is likely because of the added variations or complexities (e.g., in flower morphology and chemical constituents, and pest infections in leaves) that could not be well modeled by the rLDA classifier using existing datasets. Upon examination of the corresponding confusion matrices (**Figure 12**), the misclassification between the cultivars BaOx (BX) and Cherry Wine (CW) contributed the most to the overall accuracy deterioration, while comparable, noticeably higher accuracies were obtained for the other three cultivars. Although modeling the leaf or flower samples alone at similar growth stages led to better accuracy in classifying hemp cultivars, it would be desirable to have models that are robust to variations associated with plant organs and growth stages. Hence it is worthy of further investigations to exploit more advanced pattern classification algorithms, on a larger, more diverse set of hemp samples, to improve the accuracy of cultivar classification regardless of growth stages or plant organs.

Overall, these classification results demonstrate that hyperspectral imaging coupled with supervised modeling is a viable means for differentiating hemp cultivars with high accuracy, and that the growth stage and plant organ need to be factored in developing cultivar classification models.

Furthermore, rLDA models were built for discriminating the five growth stages and plant organs (leaf and flower) for each of the five hemp cultivars. The classification accuracies of 100% with zero standard error in 30 modeling replications (**Figure 13**) were obtained in all the scenarios. The superior results are not unexpected given the clear separation of different categories observed in the PCA space (**Figure 8**). The results are also in good agreement with the findings in literature. Borille et al. (2017) achieved 100% accuracy in discriminating three growth stages of *Cannabis sativa* using NIRS combined with support vector machine. Holmes et al. (2020) applied hyperspectral imaging in 900–1700 nm for discriminating flowers, stems and leaves of *Cannabis sativa* and achieved near 100% precision based on decision tree modeling. All these findings conceivably verify the prowess of hyperspectral imaging for accurately discriminating plant growth stages and organs (leaf and flower). Moreover, the perfect classification of growth stages can be potentially beneficial for improving the classification of hemp cultivars at varying growth stages by deploying cascade classifiers for classifying both hemp growth stages and cultivars.

Discussion

It is important to point out potential areas for further improvements. Although the standard reference (2" × 2" in size) was scanned along with hemp samples for spectral correction, it was not still sufficient for accounting for the spatial non-uniformity of illumination over the scanning line. It is more desirable to use a larger reference for imaging so that the spectral correction of samples can be performed at a pixel level along the scanning line. An alternative solution is to improve the lighting design to provide uniform illumination over samples. Using two line-light illuminators positioned symmetrically to the camera axis (Ariana and Lu, 2010), instead of the four-lamp setup in this study (**Figure 2**), may improve the illumination uniformity. This research and other previous studies on hyperspectral imaging for cannabis plants or hemp (Duchateau et al., 2020; Holmes et al., 2020; Pereira et al., 2020) did not consider spatial or textural features for modeling. Arguably these features are also useful for plant classification tasks such as cultivar differentiation, since different cultivars of hemp leaves and flowers are likely to have different morphological features, regardless of growth stages, based on which experienced agronomic experts tell apart different crop cultivars. Further research is hence warranted to extract textural features and exploit strategies of fusing them with spectral features for improving the differentiation of hemp cultivars. Meanwhile, wavelength selection or dimension reduction (e.g., PCA) can be conducted to facilitate texture feature extraction.

The present study only conducted hyperspectral scanning for sampled plant organs (e.g., leaves and flowers) in controlled-light settings. For high-throughput testing and a further validation of the hyperspectral imaging approach, further studies are needed to perform *in-situ* scanning of hemp plants under natural light conditions, requiring no sampling of plant parts. Harvesting hemp for maximum CBD yield while avoiding THC exceeding legal thresholds requires quantification of these chemical compounds in plant organs (especially flowers). Investigations are underway to determine the feasibility of using hyperspectral

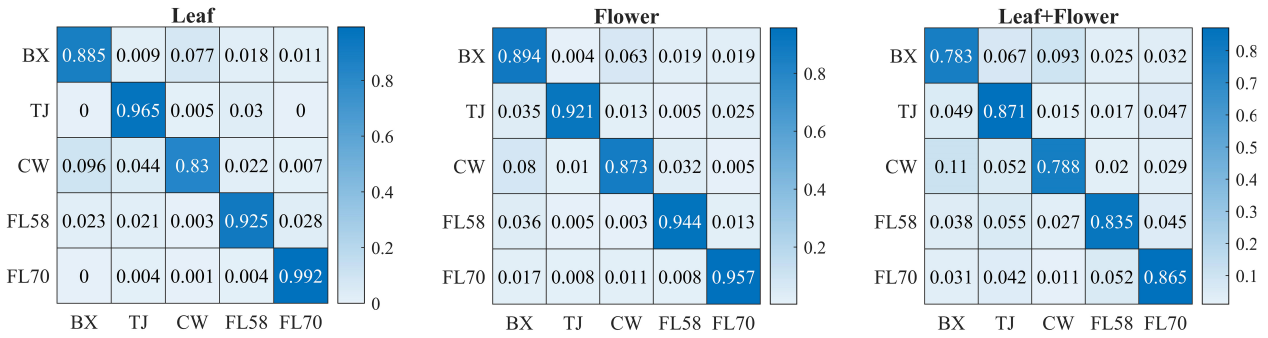


FIGURE 12 | Confusion matrices for hemp cultivar classification using the samples from all the growth stages. Each confusion matrix is obtained by pooling and row-wise normalizing classification results over 30 modeling replications.

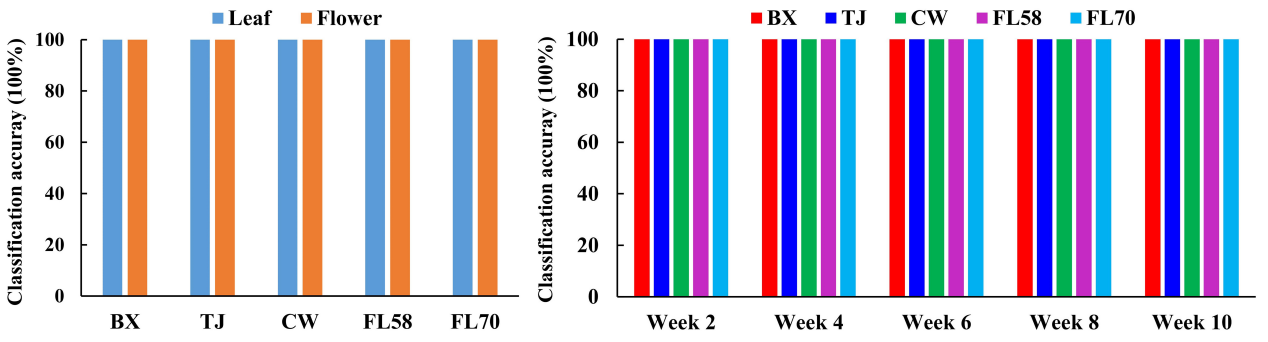


FIGURE 13 | Classification accuracies in differentiating plant growth stages **(left)** for each hemp cultivar (i.e., BX, TJ, CW, FL58, and FL70) and plant organs **(right)** at each growth stage. 100% accuracy is obtained in all the modeling scenarios with zero standard deviation for 30 replications.

imaging for screening hemp genotypes based on CBD and THC concentrations in plant tissues at different growth stages.

CONCLUSION

In this study we propose a new methodology of using hyperspectral imaging for differentiating cultivars, growth stages, and plant organs (leaves and flowers) of CBD hemp. Fresh leaves and flowers of five hemp cultivars, harvested at five growth stages 2–10 weeks after flower initiation, were scanned by a benchtop hyperspectral reflectance imaging system in the wavelength range of 400–1000 nm. An image processing algorithm was developed for segmenting samples from background. The spectral profiles and PC score scatter plots of hemp samples, to a varying degree, revealed the separation among the hemp cultivars, growth stages and plant organs. The rLDA models, using leaf or flower samples at individual growth stages, achieved the classification accuracies of 96.8%-99.6% in the differentiation of hemp cultivars. Pooling leaf and flower samples at all growth stages resulted in deteriorated accuracies compared to modeling samples at individual growth stages. Both growth stages and plant organs need to be factored in model development for hemp cultivar classification. In contrast, in the differentiation of growth stages and plant organs, the rLDA models achieved 100% accuracies consistently. This study shows that hyperspectral imaging can be used for non-destructive and accurate differentiation between hemp cultivars, growth stages and plant organs, and it is a potentially valuable tool for phenotyping, cultivar selection and optimization of harvest timing in CBD hemp production. Extensive research is still needed to develop and deploy hyperspectral imaging technology for field-scale, *in-situ* applications.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YL: conceptualization, methodology, data collection, final analysis, writing – original draft preparation, reviewing, and editing. SY: data collection, writing, reviewing, editing, and project administration. EL: data collection, writing, reviewing, and editing. BW: resources, writing, reviewing, and editing. DS: data collection, resources, writing, reviewing, and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported through funding from the North Carolina Agricultural Foundation, Inc., and the USDA National Institute of Food and Agriculture Hatch project 1021499.

ACKNOWLEDGMENTS

We would like to acknowledge Mr. Ryan Patterson for generously donating the cultivar cuttings used in these trials and Mr. Piyush Pandey for assisting in the collection of hyperspectral data.

REFERENCES

Adesina, I., Bhowmik, A., Sharma, H., and Shahbazi, A. (2020). A review on the current state of knowledge of growing conditions, agronomic soil health practices and utilities of hemp in the United States. *Agriculture* 10:129. doi: 10.3390/agriculture10040129

Ariana, D. P., and Lu, R. (2010). Evaluation of internal defect and surface color of whole pickles using hyperspectral imaging. *J. Food Eng.* 96, 583–590.

Borille, B. T., Marcelo, M. C. A., Ortiz, R. S., de Cássia Mariotti, K., Ferrão, M. F., and Limberger, R. P. (2017). Near infrared spectroscopy combined with chemometrics for growth stage classification of cannabis cultivated in a greenhouse from seized seeds. *Spectrochim. Acta A* 173, 318–323. doi: 10.1016/j.jfoodeng.2009.09.005

Brightfield Group (2019). *From to Aisle: U.S. CBD Market 2019 Report.* Available online at: https://content.brightfieldgroup.com/2019-us-cbd-market

Campbell, B. J., Berrada, A. F., Hudalla, C., Amaducci, S., and McKay, J. K. (2019). Genotype × environment interactions of industrial hemp cultivars highlight diverse responses to environmental factors. *Agrosyst. Geosci. Environ.* 2:180057. doi: 10.2134/age2018.11.0057

Capriotti, A. L., Cannazza, G., Catani, M., Cavaliere, C., Cavazzini, A., Cerrato, A., et al. (2021). Recent applications of mass spectrometry for the characterization of cannabis and hemp phytocannabinoids: from targeted to untargeted analysis. *J. Chromatogr.* 1655:462492. doi: 10.1016/j.chroma.2021.462492

Carpenter, C., and Peroutek, T. (2019). *Industrial Hemp: Overview of Opportunities and Risks*. Denver, CO: CoBank ACB.

Cherney, J. H., and Small, E. (2016). Industrial hemp in North America: production, politics and potential. *Agronomy* 6:58. doi: 10.3390/agronomy6040058

Cirrincione, M., Saladini, B., Brighenti, V., Salamone, S., Mandrioli, R., Pollastro, F., et al. (2021). Discriminating different *Cannabis sativa* L. chemotypes usingattenuated total reflectance – infrared (ATR-FTIR) spectroscopy: a proof of concept. *J. Pharm. Biomed. Anal.* 204:114270. doi: 10.1016/j.jpba.2021.114270

Cozzolino, D., Smyth, H. E., and Gishen, M. (2003). Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *J. Agric. Food Chem.* 51, 7703–7708. doi: 10.1021/jf034959s

De Backer, B., Maebe, K., Verstraete, A. G., and Charlier, C. (2012). Evolution of the content of THC and other major cannabinoids in dug-type cannabis cuttings and seedlings during growth of plants. *Am. Acad. Forensic Sci.* 57, 918–922. doi: 10.1111/j.1556-4029.2012.02068.x

Dingha, B., Sandler, L., Bhowmik, A., Akotsen-Mensah, C., Jackai, L., Gibson, K., et al. (2019). Industrial hemp knowledge and interest among North Carolina organic farmers in the United States. *Sustainability* 11:2691. doi: 10.3390/su11092691

Dong, W., Liang, J., Barnett, I., Kline, P. C., Altman, E., and Zhang, M. (2019). The classification of *Cannabis* hemp cultivars by thermal desorption direct analysis in real time mass spectrometry (TD-DART-MS with chemometrics. *Anal. Bioanal. Chem.* 411, 8133–8142. doi: 10.1007/s00216-019-02200-7

Duchateau, C., Kauffmann, J. K., Canfyn, M., Stevigny, C., Braekeleer, K. D., and Deconinck, E. (2020). Discrimination of legal and illegal *Cannabis* spp. According to European legislation using near infrared spectroscopy and chemometrics. *Drug Test Anal.* 12, 1309–1319. doi: 10.1002/dta.2865

Ferwerda, J. G., Skidmore, A. K., and Mutanga, O. (2005). Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. *Int. J. Remote Sens.* 26, 4083–4095. doi: 10.1080/01431160500181044

Glasbey, C. A. (1993). An analysis of histogram-based thresholding algorithm. *CVGIP Graph. Model. Image Process.* 55, 532–537. doi: 10.1006/cgip.1993.1040

Glivar, T., Erzen, J., Kreft, S., Zagozen, M., Cerenak, A., Ceh, B., et al. (2020). Cannabinoid content in industrial hemp (*Cannabis sativa* L.) varieties grown in Slovenia. *Ind. Crops Prod.* 145:112082. doi: 10.1016/j.indcrop.2019.112082

Guo, Y., Hastie, T., and Tibshirani, R. (2007). Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* 8, 86–100. doi: 10.1093/biostatistics/kxj035

Holmes, W. S., Ooi, M. P. L., Kuang, Y. C., Simpkin, R., Lopez-Ubiria, I., Vidiella, A., et al. (2020). "Classifying *Cannabis sativa* flowers, stems and leaves using statistical machine learning with near-infrared hyperspectral reflectance imaging," in *Proceedings of the 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, (Dubrovnik), 1–6. doi: 10.1109/I2MTC43012.2020.9129531

Horler, D. N. H., Dockray, M., and Barber, J. (1983). The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4, 273–288. doi: 10.1080/01431168308948546

Hunt, E. R., and Rock, B. N. (1989). Detection of changes in leaf water content using near-and middle-infrared reflectances. *Remote Sens. Environ.* 30, 43–54. doi: 10.1016/0034-4257(89)90046-1

Jin, D., Jin, S., Yu, Y., Lee, C., and Chen, J. (2017). Classification of *Cannabis* cultivars marketed in Canada for medical purposes by quantification of cannabinoids and terpenes using HPLC-DAD and GC-MS. *J. Anal. Bioanal. Tech.* 8:1000347. doi: 10.4172/2155-9872.1000349

Lu, Y., and Lu, R. (2017). Histogram based automatic thresholding for bruise detection of apples by structured-illumination reflectance imaging. *Biosyst. Eng.* 160, 30–41. doi: 10.1016/j.biosystemseng.2017.05.005

Lu, Y., Du, C., Yu, C., and Zhou, J. (2014). Classifying rapeseed varieties using Fourier transform infrared photoacoustic spectroscopy (FTIR-PAS). *Comput. Electron. Agric.* 107, 58–63. doi: 10.1016/j.compag.2014.06.005

Lu, Y., Payn, K. G., Pandey, P., Acosta, J. J., Heine, A. J., Walker, T. D., et al. (2021). Hyperspectral imaging with cost-sensitive learning for high-throughput screening of loblolly pine (*Pinus taeda* L.) seedling for freeze tolerance. *Trans. ASABE* 64, 2045–2059. doi: 10.13031/trans.14708

Lu, Y., Saeys, W., Kim, S., Peng, Y., and Lu, R. (2020). Hyperspectral imaging technology for quality and safety evaluation of horticultural products: a review and celebration of the past 20-year progress. *Postharvest Biol. Technol.* 170:111318. doi: 10.1016/j.postharvbio.2020.111318

Luo, W., Huan, S., Fu, H., Wen, G., Cheng, H., Zhou, J., et al. (2011). Preliminary study on the application of near infrared spectroscopy and pattern recognition methods to classify different types of apple samples. *Food Chem.* 128, 555–561. doi: 10.1016/j.foodchem.2011.03.065

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., and Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* 106, 135–141. doi: 10.1034/j.1399-3054.1999.106119.x

Pereira, J. F. Q., Pimentel, M. F., Amigo, J. M., and Honorato, R. S. (2020). Detection and identification of *Cannabis sativa* L. using near infrared hyperspectral imaging and machine learning methods: a feasibility study. *Spectrochim. Acta A* 237:118385. doi: 10.1016/j.saa.2020.118385

Otsu, N. (1979). A threshold selection method from gray-scale histograms. *IEEE Trans. Syst. Man Cybern.* 9, 62–66. doi: 10.1109/TSMC.1979.4310076

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv* [Preprint] arXiv:1811.12808,

Renée, J. (2018). *Hemp As An Agricultural Commodity*. Washington, DC: Congressional Research Service.

Sanchez, L., Baltensperger, D., and Kurouski, D. (2020). Raman-based differentiation of hemp, cannabidiol-rich hemp, and cannabis. *Anal. Chem.* 92, 7733–7773. doi: 10.1021/acs.analchem.0c00828

Schluttenhofer, C., and Yuan, L. (2019). Hemp hemp hooray for *Cannabis* research. *Science* 363, 701–702. doi: 10.1126/science.aaw3537

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *arXiv* [Preprint] arXiv:1206.2944,

Stack, G. M., Toth, J. A., Carlson, C. H., Cala, A. R., Marrero-González, M. I., Wilk, R. L., et al. (2021). Season-long characterization of high-cannabinoid hemp (*Cannabis sativa* L.) reveals variation in cannabinoid accumulation, flowering time, and disease resistance. *GCB Bioenergy* 13, 546–561. doi: 10.1111/gcbb.12793

USDA-AMS (2021). *Establishment of A Domestic Hemp Production Program*. Washington, DC: Agricultural Marketing Service, U.S. Department of Agriculture.

Wang, X., Harrington, P. B., and Baugh, S. F. (2018). Effect of preprocessing high resolution mass spectra on the pattern recognition of *Cannabis*, hemp, and liquor. *Talanta* 180, 229–238. doi: 10.1016/j.talanta.2017.12.032

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 February 2022
doi: 10.3389/fpls.2022.834938

Wheat-Net: An Automatic Dense Wheat Spike Segmentation Method Based on an Optimized Hybrid Task Cascade Model

Jiajing Zhang[1,2,3], An Min[4], Brian J. Steffenson[5], Wen-Hao Su[6], Cory D. Hirsch[5], James Anderson[7], Jian Wei[1], Qin Ma[1] and Ce Yang[4]**

[1] College of Information and Electrical Engineering, China Agricultural University, Beijing, China, [2] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [3] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China, [4] Department of Bioproducts and Biosystems Engineering, University of Minnesota, Saint Paul, MN, United States, [5] Department of Plant Pathology, University of Minnesota, Saint Paul, MN, United States, [6] College of Engineering, China Agricultural University, Beijing, China, [7] Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN, United States

OPEN ACCESS

Edited by:
Dongjian He,
Northwest A&F University, China

Reviewed by:
Yong Suk Chung,
Jeju National University, South Korea
Huijun Yang,
Northwest A&F University, China
Honghua Jiang,
Shandong Agricultural University, China

***Correspondence:**
Qin Ma
sockline@163.com
Ce Yang
ceyang@umn.edu

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: 14 December 2021
Accepted: 18 January 2022
Published: 10 February 2022

Citation:
Zhang J, Min A, Steffenson BJ, Su W-H, Hirsch CD, Anderson J, Wei J, Ma Q and Yang C (2022) Wheat-Net: An Automatic Dense Wheat Spike Segmentation Method Based on an Optimized Hybrid Task Cascade Model. Front. Plant Sci. 13:834938. doi: 10.3389/fpls.2022.834938

Precise segmentation of wheat spikes from a complex background is necessary for obtaining image-based phenotypic information of wheat traits such as yield estimation and spike morphology. A new instance segmentation method based on a Hybrid Task Cascade model was proposed to solve the wheat spike detection problem with improved detection results. In this study, wheat images were collected from fields where the environment varied both spatially and temporally. Res2Net50 was adopted as a backbone network, combined with multi-scale training, deformable convolutional networks, and Generic ROI Extractor for rich feature learning. The proposed methods were trained and validated, and the average precision (AP) obtained for the bounding box and mask was 0.904 and 0.907, respectively, and the accuracy for wheat spike counting was 99.29%. Comprehensive empirical analyses revealed that our method (Wheat-Net) performed well on challenging field-based datasets with mixed qualities, particularly those with various backgrounds and wheat spike adjacence/occlusion. These results provide evidence for dense wheat spike detection capabilities with masking, which is useful for not only wheat yield estimation but also spike morphology assessments.

Keywords: wheat spike, instance segmentation, Hybrid Task Cascade model, challenging dataset, non-structural field

INTRODUCTION

Wheat is the most widely cultivated cereal crop and also one of the most important food sources for humans in the world. The spike is the most important component of the wheat plant because it contains the seeds that are harvested and ultimately consumed. Therefore, in-field automated wheat spike detection based on remote sensing is an important step toward yield estimation and spike morphology assessments.

To detect wheat spikes, the remote sensing imaging devices are useful tools to replace traditional artificial detection (Aparicio et al., 2000). Hyperspectral imaging cameras can provide rich spectral information for wheat detection (Shen et al., 2021), but the cost of hyperspectral imaging is expensive, which restricts the application in various fields (Zhang et al., 2020). Thus, the cheaper RGB imaging camera is a realistic alternative to achieve effective wheat detection. Deep learning (DL) with strong feature learning abilities has spawned a multitude of applications in RGB images. It encodes the composition of lower-level features into more discriminative higher-level features (Ni et al., 2019). DL can solve more complex problems with higher precision and has been successfully used in plant classification (Sun et al., 2019; Yang et al., 2019; Khaki et al., 2020), yield prediction (Pound et al., 2017; de Luna et al., 2020; Zhuang et al., 2020), growth monitoring (Kovalchuk et al., 2017; Qiongyan et al., 2017), and disease/pest detection (Senthilkumar et al., 2017; da Silva et al., 2019; Desai et al., 2019). Thus, DL, with its advantages of high precision and intelligence, is an attractive alternative to conventional wheat spike detection methods (Germain et al., 1995; Cointault et al., 2008a,b).

Recently, DL has been shown to perform well in a wide variety of wheat spike detection studies. Some previous works involving wheat detection have been conducted under laboratory conditions and controlled environments (Hasan et al., 2018; Sadeghi-Tehran et al., 2019; Chandra et al., 2020; Misra et al., 2020). Laboratory-based experiments have good lighting conditions and a clean background, which is not the case for field-based research, which is more complicated and yields images where the background usually contains a lot of disturbances (including soil and weeds). The complicated background greatly increases the difficulty of resolving individual wheat spikes but represents the actual growing environment of wheat. Thus, models developed from the field are more realistic of real-world conditions for wheat cultivation. Several in-field spike detection and counting studies have been conducted (David et al., 2020, 2021; Xu et al., 2020; Wang et al., 2021). Among them, David et al. (2021) constructed a more diverse and less noisy Global Wheat Head Detection (GWHD) dataset, which promoted the development of wheat spike detection. The detection results from these studies were based on bounding boxes, which can be used for counting wheat spikes. However, the precise pixel areas of wheat are often required in wheat management (such as evaluation of spikes disease and accurate yield prediction), which cannot be achieved by detecting the bounding box only by segmentation. Segmentation provides information such as size, shape, and relative location of the segments in the image, which can be used for phenotypic traits such as spike size, shape, distribution, and wheat yield potential. Therefore, it is necessary to explore an approach of segmenting wheat spikes to meet the needs for precise spike areas in wheat management.

There are some researchers who have used semantic segmentation algorithms to segment wheat spikes in the field with a simpler environment by controlling some factors in the experiment. For example, in implementing a Fully Convolutional Network (FCN) segmentation model of individual wheat spikes, Zhang et al. (2019) positioned spikes to avoid occlusion–an intervention that does not simulate the actual growing conditions of wheat in the field. Alkhudaydi et al. (2019) employed FCN to segment multiple wheat spikes, which achieved a Mean Accuracy (MA) of classification of $> 76\%$. However, their model performed poorly under challenging conditions caused by variable lighting and weather (Alkhudaydi et al., 2019). Tan et al. (2020) performed simple linear iterative clustering (SLIC) for superpixel segmentation of digital wheat images, which resulted in a high accuracy (94.01%) under high nitrogen fertilizer level and a lower accuracy (80.8%) under no nitrogen fertilizer application. Ma et al. (2020) developed EarSegNet to segment multiple wheat spikes from canopy images captured under field conditions and realized a precision of 79.41%. However, semantic segmentation algorithms cannot segment wheat spikes out individually when they are obstructed by other spikes, which is a common situation under field conditions.

Instance segmentation can effectively segment partially obstructed wheat spikes. This method localizes objects of interest in an image at the pixel level, which achieves both object detection and semantic segmentation (Li et al., 2017; Chen et al., 2019). With instance segmentation, the segmented objects are generally represented by masks and a bounding box (bbox); however, few studies have been advanced using instance segmentation for detecting wheat spikes under field settings. Qiu et al. (2019) used a Mask RCNN model to reliably detect wheat spikes (mean average precision is 0.9201) with different shapes and features in the field. However, to achieve these results, they used a background plate to block complex backgrounds and also a shade shed to provide even lighting, which reduced the complexities of image capture and subsequent annotation. They also divided the original image into many smaller images, which resulted in image distortion. This, in turn, resulted in images with only partial objects or no objects at all, which would destroy the integrity of the wheat spikes. In our previous research (Su et al., 2021), we basically realized the instance segmentation of wheat in a complex field environment, but its low accuracy cannot meet the needs of practical applications and further research is necessary to achieve high-precision instance segmentation.

In summary, the object detection of wheat is insufficient for accurate phenotype study and semantic segmentation cannot segment common occlusive wheat spikes. Instance segmentation methods can solve the above problems, but the conventional instance segmentation methods of wheat are either in laboratory conditions or controlled environments or have low accuracy, etc., which may not be suitable for phenotyping spikes under complex field environment. Therefore, it is necessary to explore a more applicable and accurate approach for segmenting wheat spikes under field conditions. Therefore, the specific objectives of this study were to: construct a new instance segmentation model (called Wheat-Net) based on a multi-task Hybrid Task Cascade (HTC) model (Chen et al., 2019) that can precisely instance segment wheat spikes in high densities in the field. Comprehensive empirical analyses reveal that Wheat-Net achieved excellent performance on a challenging dataset with various complex backgrounds and a high level of obstruction. In a complex, unstructured environment, our method not only accurately detected the wheat spikes with

bounding boxes but also extracted spike regions from the background at the pixel level.

MATERIALS AND METHODS

Data Collection

Wheat genotypes were sown in field plots on the St. Paul campus of the University of Minnesota (UMN) in 2019. These genotypes included mostly breeding lines from the UMN hard red spring wheat breeding program, which can vary for different spike morphology traits such as color, shape types as well as spike density. The images at the late flowering stage (July 11) to the milk stage of maturity (August 2) were collected from the field including 20 wheat genotypes, which can enhance the adaptability of the model to different wheat genotypes. In the complex field environment, we used the camera of Canon EOS Rebel T7i (autofocus single-lens reflex, pixels: 6,000 × 4,000) to collect image data under different weather conditions (including sunny and cloudy days). The exposure time, white balance, and ISO speed were automatically set based on the automatic mode of camera. The distance from the object is about 1–2 m. The wheat images collected had complex backgrounds, including weeds, soil, blurred wheat, blue sky, and white clouds. We expect that users take images of the wheat/barley trial plots with very loose image acquisition requirements (e.g., imaging angle and distance). Therefore, we acquired the current dataset with various angles and distances, which also can increase the diversity of data and enhance the adaptability and robustness of the model.

Wheat is typically a dense crop and the images (**Figure 1A**) collected contained as many as 124 spikes per image. In addition, it was common that portions of images had insufficient illumination (blue box with zoom-in shown in **Figure 1B**). Moreover, the above factors also resulted in many problems such as spike adjacence (Red box with zoom-in shown in **Figure 1C**), occlusion, variation in spike size, and partial spikes on the image edge (Yellow box with zoom-in shown in **Figure 1D**). The spike occlusion problem was the most serious problem and included various scenarios such as spikes over spikes (**Figure 2A**), leaves over spikes (**Figure 2B**), stems over spikes (**Figure 2C**), and awns over spikes (**Figure 2D**). Although the above factors greatly increase the segmentation difficulty, they encompass the true field environment and are helpful to improve the robustness of the spike segmentation model.

The high complexity of the images brings great challenges to artificial annotation. The artificial image annotation software, Labelme (Russell et al., 2008), was used to label the ground truth for wheat spikes using polygons. **Figure 3** shows the annotation of the images in this paper. To obtain high-quality annotated datasets, we enlarged the image about 200% or larger and selected the dense points along the outside edge of every spike to form an accurate spike region. However, there are still several very blurred spikes in the enlarged picture, which cannot be distinguished by the humans and are not annotated. Our group put a lot of effort in annotation and we believe this dataset can promote further wheat phenotypic studies. These annotated images were used to calculate the loss and optimize the model parameters during model training. In machine learning, about 2/3 to 4/5 of the datasets are usually used for training, and the remaining images are used for testing. Therefore, there are 524 images in the training set (12,591 spikes) and 166 images in the test set (4,934 spikes) in this paper.

Methods of Wheat Instance Segmentation

Architecture of Wheat-Net

In this study, instance segmentation was the key protocol implemented to reliably detect and segment wheat spikes in a complex non-structural environment. We built the wheat spike instance segmentation model, Wheat-Net, for our high-complexity dataset based on the HTC model (Chen et al., 2019), which is a novel cascade architecture for instance segmentation. The HTC model has a powerful cascade structure that enhanced performance on various tasks. It solved the problem of insufficient information flow between mask branches at different stages in Cascade Mask RCNN, which is a direct combination of Cascade RCNN (Cai and Vasconcelos, 2021) and Mask RCNN (He et al., 2017). The HTC model effectively integrated cascade into instance segmentation by interweaving detection and segmentation for joint multi-stage processing, achieving outstanding performance on COCO (Common Objects in

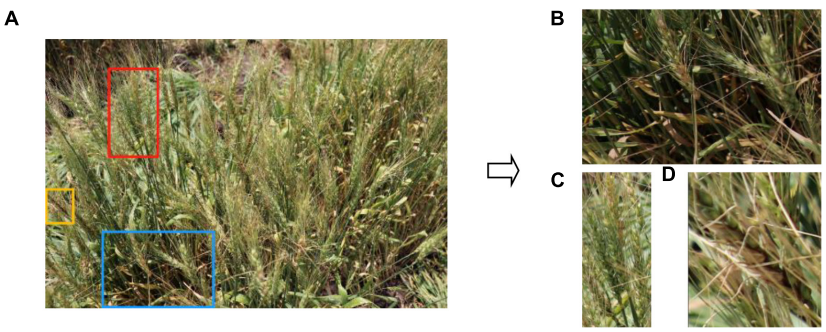


FIGURE 1 | (A) An example of an original image of a wheat plot indicating sections (blue, red, and yellow boxes) enlarged to show **(B)** an area with incomplete illumination, **(C)** adjacent spikes in close proximity, and **(D)** partial spikes on the edge of the images.

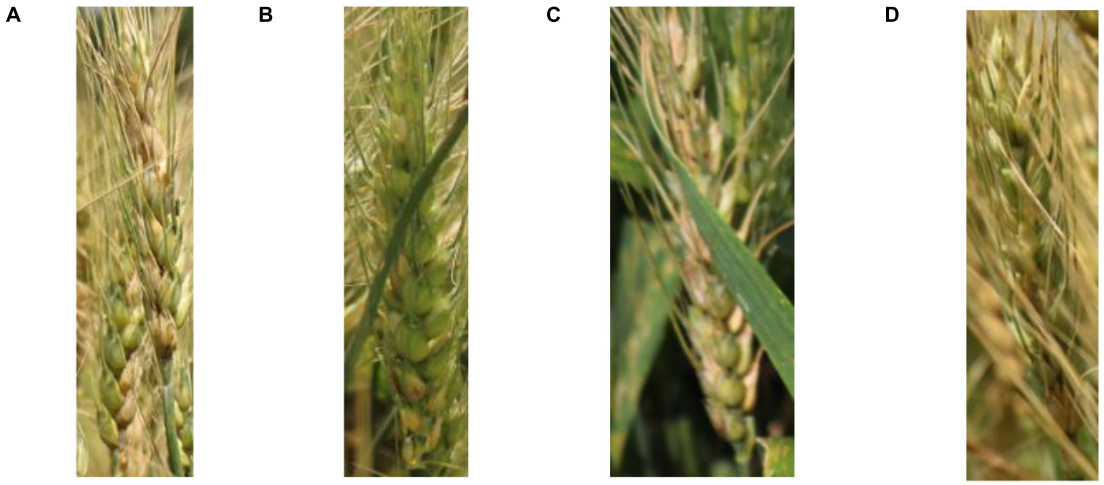


FIGURE 2 | Examples of various spike occlusion scenarios: **(A)** spike over spike, **(B)** leaf over spike, **(C)** stem over spike, and **(D)** awns over spike.

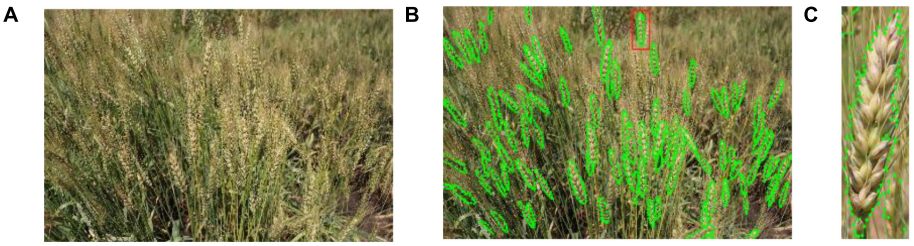


FIGURE 3 | Annotation of wheat spikes: **(A)** the original image of a wheat plot, **(B)** the image with annotated wheat spikes, and **(C)** details of a single annotated spike.

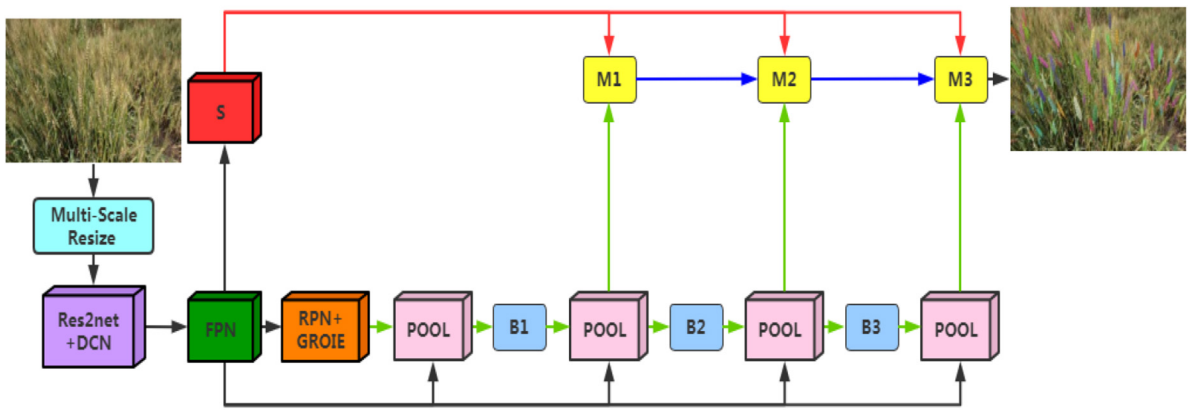


FIGURE 4 | The architectures of Wheat-Net. "POOL" region-wise feature extraction, "B" bounding box, and "M" mask. "S" is semantic segmentation branch.

Context) test-dev and test-challenge (Lin et al., 2014). We cascaded three Mask RCNN networks to build the Wheat-Net (**Figure 4**). The advantages of this model can be ascribed to three key aspects. (1) It interleaved the box and mask branches (the green lines in **Figure 4**) based on Cascade Mask RCNN. This improvement allowed the mask branch to take advantage of the updated bbox. For instance segmentation of wheat spikes, the bbox information is very important for wheat mask segmentation. If bbox detects two adjacent spikes as the same object, the model will difficult to segment them. Therefore, the interleaving of box and mask branches can help to achieve more accurate wheat spike segmentation. (2) It made full use of the mask feature of the preceding stage by adding a direct information flow between mask branches (the blue lines in

Figure 4). The direct information flow can learn more abundant multi-scale information of wheat from complex images, which further improved the accuracy of wheat segmentation. (3) It explored more contextual information by adding a semantic segmentation branch (the red lines in **Figure 4**), which can help the wheat spikes to be segmented accurately from the complex background. The above optimizations are combined (Equations 1–5) for better predictions, which effectively improved the utilization of information and enhanced performance.

$$r_t = B_t(x_t^{box}) \tag{1}$$

$$x_t^{box} = p\left(x, r_{t-1}\right) + p(S\left(x\right), r_{t-1}) \tag{2}$$

$$x_t^{mask} = p\left(x, r_t\right) + p(S\left(x\right), r_t) \tag{3}$$

$$\mathbf{m_t} = \mathbf{M_t}(\mathcal{F}(\mathbf{x_t^{mask}}, \mathbf{m_{t-1}^{-}})) \tag{4}$$

$$\mathcal{F}\left(\mathbf{x_t^{mask}}, \mathbf{m_{t-1}^{-}}\right) = \mathbf{x_t^{mask}} + \mathbf{g_t}(\mathbf{m_{t-1}^{-}}) \tag{5}$$

Where x is the feature of the backbone network, x_t^{box} and x_t^{mask} denote box and mask features of x and the input Region of Interest (RoI). S indicates the semantic segmentation head. The box and mask heads of each stage take the RoI features extracted from the backbone as input. $p\ (\cdot)$ is a pooling operator, B_t and M_t indicate the box and mask head at the t-th stage.r_t and m_t represent predictions of box and mask, respectively. m_{t-1}^{-} indicates the intermediate feature of M_{t-1}. $\mathcal{F}$ is a function that combines the features of the current stage and the preceding one. g_t denotes a 1×1 convolutional layer.

Optimization of Wheat-Net

Different backbones have an important effect on the performance of the model because of their differences in feature extraction ability. Res2Net50 (Gao et al., 2021; **Figure 5B**) represents multi-scale features at a granular level and increases the range of receptive fields for each network layer, which is different from the concurrent bottleneck structure shown in **Figure 5A**, such as ResNet (He et al., 2016). Specifically, it replaces the 3×3 filters of n channels with a set of smaller filter groups, which are connected in a hierarchical residual-like style to increase the number of scales that the output features can represent. It can capture more details and global features of wheat without increasing calculations for wheat segmentation. ResNeXt (Xie et al., 2017) is an improved model of ResNet (**Figure 5C**), and is constructed by repeating a building block and the transformations to be aggregated, all of the same topology. It is a simple, homogeneous, and multi-branch architecture, which can extend to any large number of transformations without specialized designs. In the experimental part of this paper, we compare the performance of the above-mentioned backbones in our dataset.

Deformable convolutional networks (DCN) (Dai et al., 2017) were integrated into our model because they provide a solution to model dense spatial transformations and are effective for sophisticated vision tasks. DCN allowed free deformation of the sampling grid as shown in **Figure 6**, which added offsets learned from target tasks to the regular sampling grid of standard convolution without additional supervision. DCN can help to solve the geometric deformation and enhance the robustness of the model for segmenting various sizes and angles of wheat spikes.

In our model, feature pyramid networks (FPN) (He et al., 2017) extracted RoI features from different levels of the feature pyramid by using a top-down architecture. These different features, generated and fused by FPN, comprised the inputs of the Region Proposal Network (RPN) (Ren et al., 2017). RPN predicted object bounds and objectness scores to efficiently generate region proposals with a wide range of scales and aspect ratios. Generic RoI Extractor (GRoIE) (Rossi et al., 2020) was used to extract the RoI. Since all layers of FPN retain useful information of wheat spikes, non-local building blocks and attention mechanisms were introduced to extract more information of wheat and overcome the limitations of existing RoI extractors, which select only one (the best) layer from FPN. They also can be integrated seamlessly with the two-stage architectures for instance segmentation tasks for superior performance compared to traditional RoI extractors (Pont-Tuset et al., 2017). Multi-task learning (Caruana, 1997) combined all tasks into a single model: that is, what is learned for each task can help other tasks be learned better. In this paper, we used multi-task learning to achieve both target detection and semantic segmentation of wheat spikes. Hence, Multi-task learning can improve learning efficiency and prediction accuracy by learning multiple objectives from a shared representation.

As an important part of the object detection pipeline, non-maximum suppression (NMS) could sort the detection bbox based on their scores (Rosenfeld and Thurston, 1971), select the detection bbox with the highest score and suppress all other bbox that had significant overlap (using a predefined threshold) with it. However, NMS might lose the objects that are within the predefined overlap threshold. Due to wheat is dense plant, there are a lot of overlap wheat spikes in the images. NMS might only detect one spike between two overlap spikes. Bodla et al. (2017) proposed a Soft-NMS algorithm to prevent objects from being eliminated. It decayed the detection scores of all other objects as a continuous function of their overlap. In the experimental part of this paper, we conducted a comparative experiment between NMS and Soft-NMS.

The size of the input image had a significant impact on model performance. Because we collected images with various shooting distances and angles, the dataset contained many small spikes. In our paper, small and big spikes are labeled as ground truth, which is more in line with the actual field of wheat. Because the feature map generated by the network was much smaller than the original image, the model may lose features of small spikes and unable to detect small spikes. Therefore, if the model fails to detect small wheat spikes, the performance of the model will be affected. Multi-scale training (He et al., 2015), which defines several fixed scales in advance and randomly selects a scale for training in each epoch, can effectively improve this limitation. Therefore, we used images of multiple scales for training to improve the robustness

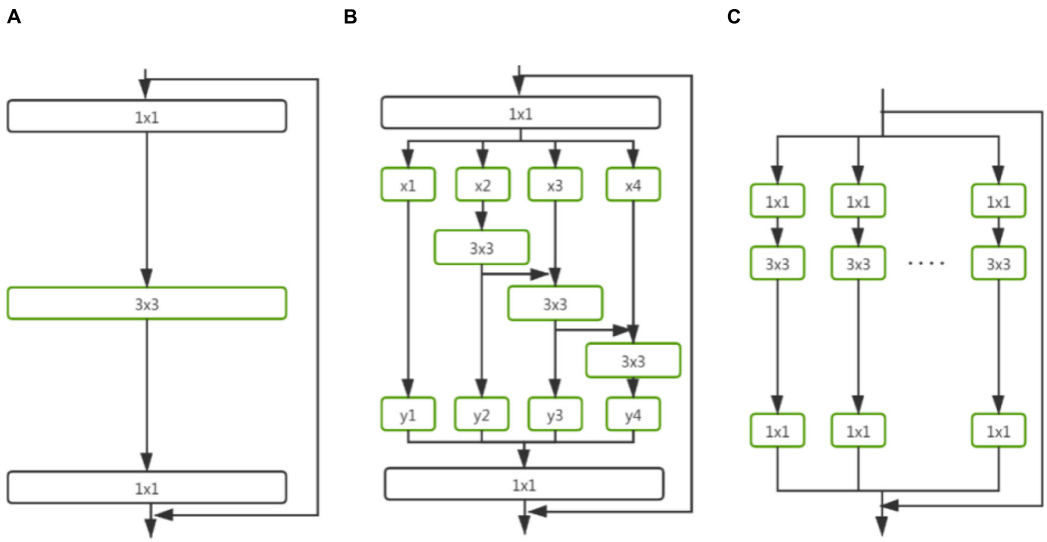


FIGURE 5 | Comparison between different backbones for Wheat-Net. **(A–C)** Are the block of ResNet, Res2Net, and ResNeXt, respectively.

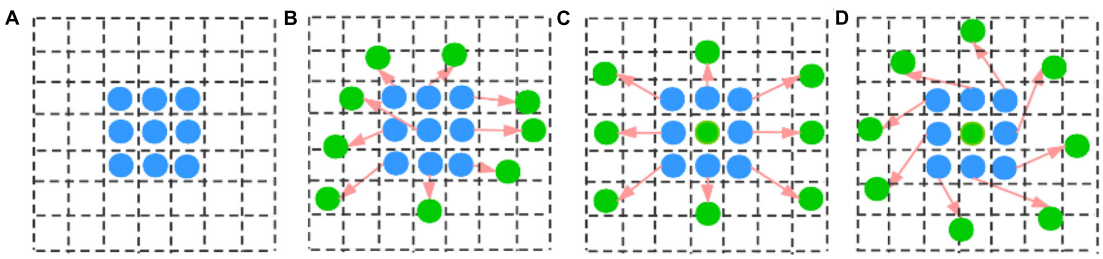


FIGURE 6 | Different calculation positions under: **(A)** standard convolution (blue points); **(B)** deformable convolution (with green points). **(C,D)** Are special cases of **(B)**, showing that the deformable convolution generalizes scale and rotation transformations.

and accuracy of our model. Due to memory constraints, the short-side of the input images was randomly selected from 416 to 1,184, and another side's size was calculated according to the aspect ratio of the original image's size.

Learning rate (LR) was one of the most important hyperparameters in training. If the LR is large at the beginning of training, the model may become unstable, making it difficult to reach the optimal solution. To address this, we used warm-up LR (He et al., 2016) to improve the training situation. Warm-up LR allows the LR to gradually increase from a small value in the first few epochs until the initial LR is reached. In this way, the model can gradually stabilize, and the convergence speed becomes faster after stabilization.

As an important hyperparameter in deep learning, LR could determine whether and when the model can converge. A large LR will make the model fluctuate greatly, and it is difficult to reach the optimal solution. In addition, as the number of iterations increases, the LR will continue to decay to reduce fluctuations of model. We chose two popular LR decay methods and compared them in the experimental chapter: one was MultiStepLR, which used the dynamic step to update the LR, and the other was CosineAnnealingLR, which decayed the LR periodically based on the cosine function. Hyperparameters of the model were adjusted and optimized based on multiple experiments. Finally, the initial LR was set to 0.0025 and adjusted every 20 epochs with a decay factor of 0.5. The other hyperparameters of the model are shown in **Table 1**.

Eventually, a new wheat spike segmentation method based on the HTC model combined with the backbone of Res2Net50, deformable convolutional networks, and Generic RoI Extractor was constructed (**Figure 4**). During the model training, each image was augmented using multiple methods (including

TABLE 1 | Hyperparameter values which optimized through training.

Parameter	Value
Optimization algorithm	SGD
Momentum	0.9
Initial learning rate	0.0025
Warmup_iterations warmup_ratio = 0.001	500
Warmup_ratio	0.001
Optimal epoch	38
Batch size	1

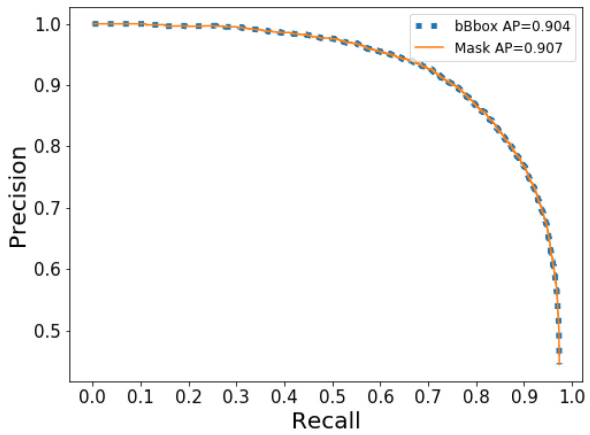


FIGURE 7 | The curve of precision and recall.

VerticalFlip, RandomBrightnessContrast, RGBShift RGB, HueSaturationValue, ChannelShuffle, Blur, and MedianBlur) and the Res2Net50 backbone was pretrained based on the ImageNet dataset (Deng et al., 2009) using transfer learning, which was suitable for solving the problem of a small training dataset. The overall loss function takes the form of multi-task learning and was defined as Equation (6).

$$L = \sum_{t=1}^{T} a_t \left(L_{bbox}^t + L_{mask}^t \right) + \beta L_{seg} \tag{6}$$

Where: L_{bbox}^t is the loss of the bounding box predictions at stage t. L_{mask}^t is the loss of mask prediction at stage t. $L_{\text{seg}}$is the semantic segmentation loss in the form of cross-entropy. Because we cascade 3 Mask RCNN networks to build the Wheat-Net architectures, T was set to 3. In addition, to balance the contributions of different stages and tasks, we set $\alpha = [1, 0.5, 0.25]$ and $\beta = 1$ by default [31].

Evaluation Metric

The performance of Wheat-Net was evaluated by average precision (AP), which is the area under the curve of precision-recall (PR) (Equations 7–9). A high AP value indicates that a model has both high precision and high recall. AP stood out as the most-used metric due to its representativeness and simplicity. AP was calculated (Equation 10) by using the method of the COCO dataset, which interpolated through all points. In this research, we evaluated the performance of Wheat-Net based on the IOU (Equation 11) threshold of 0.5, which is commonly used for instance segmentation model. The evaluation metrics are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$p_{interp}\left(r_{n+1}\right) = \underset{\tilde{r}:\tilde{r} \geq r_{n+1}}{max}\ p\left(\tilde{r}\right) \tag{9}$$

$$AP = \sum_{n=0} \left(r_{n+1} - r_n\right) p_{interp}\left(r_{n+1}\right) \tag{10}$$

$$IoU\left(A, B\right) = \left| \frac{A \bigcap B}{A \bigcup B} \right| \tag{11}$$

Where TP indicates the correct detection of wheat spikes, FP is the wrong detection of wheat spikes, and FN represents the ground truth of wheat spikes not detected. Precision indicates how many wheat spikes detected by the model are real wheat spikes. Recall indicates how many real wheat spikes are detected by model in all real spikes. p ($\tilde{r}$) is the measured precision at recall ($\tilde{r}$.) IOU is the intersection over union between two bboxes. A represents the bbox labeled manually and B represents the bbox generated based on Wheat-Net.

RESULTS

The data analysis was performed with the deep learning development framework of PyTorch. An Intel (R) Core (TM) i7-6700 processor, a 16GB random-access memory card, and a graphic card (NVIDIA GeForce GTX1080Ti 11GB) were used for the modeling process.

To determine the appropriateness of the model, the test set was used to assess the model. The AP of bbox and mask reached 0.904 and 0.907, respectively. In the case of dense wheat spike detection from complex backgrounds, false positives tended to happen more often than false negatives. Therefore, we used the PR curve (**Figure 7**), which emphasized the evaluation of the prediction model on positive examples to evaluate the performance of the model. This step confirmed the effectiveness of Wheat-Net for detecting wheat spikes in the complex field environment.

In addition, we visualized the detection results of the complex image shown in **Figure 8A**. As shown in **Figure 8**, in the non-structural field, the model showed outstanding performance for complex backgrounds, dense spikes, adjacency, and occlusion (**Figure 8B**), insufficient illumination (**Figure 8D**), and incomplete spikes on the edge of images (**Figure 8C**).

The model can effectively solve the problem of various occlusion scenarios, which is one of the most challenging areas in the field of object detection. **Figure 9** demonstrated the detection results of various occlusion scenarios including: when a spike is obstructed by another spike; when a spike is occluded by a leaf; when a spike is occluded by a stem; and when a spike is occluded by awns from another spike. Comparing the total number of spikes (4,899) detected by the model with the actual number of spikes manually labeled (4,934), 99.29% of the manually labeled wheat spikes (clearly visible to humans) are detected. It should be noted that the main goal of this paper is to accurately segment wheat spikes in complex environments, so the datasets and scenarios may different from pure wheat counting studies. This demonstrates that Wheat-Net was effective for automatic wheat spike detection under complex field conditions. In addition, the

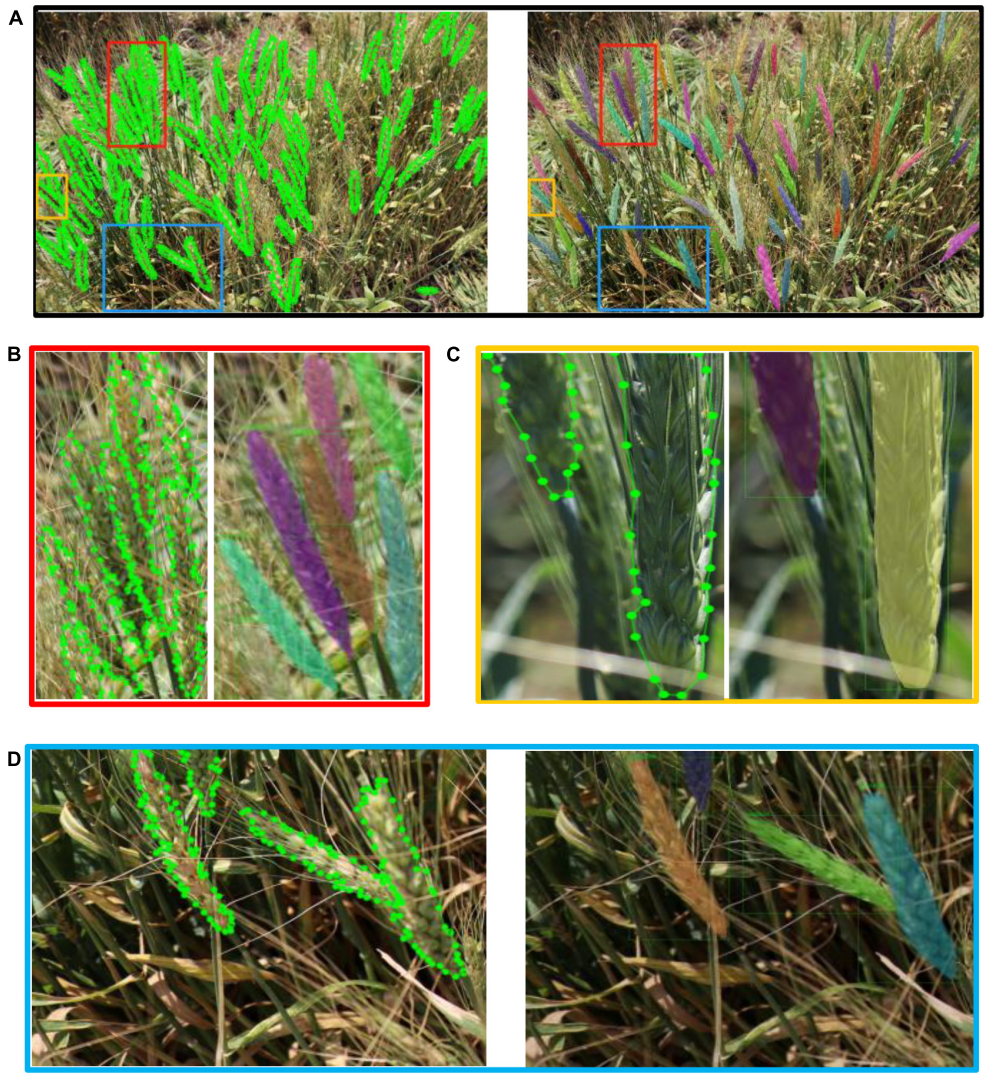


FIGURE 8 | Annotation images vs. detection results. **(A)** Overall detection results, **(B)** detail 1—area of adjacence and occlusion, **(C)** detail 2—area of incomplete spikes in image, **(D)** detail 3—area of insufficient illumination of spikes.

instant segmentation algorithm is used to segment wheat spikes out of plot images. Segmentation provides information such as size, shape, and relative location of the segments in the image, which can be used for phenotypic traits such as spike size, shape, distribution, and wheat yield potential.

Ablation Study

Ablation study is an effective way to see how a method affects the performance of the entire model by removing that specific method from the model. To perform this analysis, we used multi-scale training, DCN, and GRoIE methods to improve the performance of the model. To more accurately evaluate the effect of each method, we conducted the ablation study experiments with the Wheat-Net based on Res2net50 (LR = 0.0025, batch size = 1, image scale = 2,100*1,184) and compared the performances on the test set.

The experimental results (**Table 2**) showed that multi-scale training, DCN, and GRoIE had various effects on the performance of Wheat-Net. Specifically, the AP (both IOU = 0.5 and IOU = 0.75) were significantly improved by multi-scale training, although it increased some test times. The improvement of DCN for IOU = 0.75 was greater than that for IOU = 0.5, which showed that DCN had a more significant effect on a large IOU threshold. In addition, GRoIE increased AP with IOU = 0.5 and decreased AP with IOU = 0.75. The experimental results showed that GRoIE did not work for our dataset when using a larger threshold of IOU.

Comparative Evaluation

To make full use of the advantages of the Wheat-Net to achieve better performance, we conducted experiments to select the best backbone to build the Wheat-Net. As shown in **Table 3**,

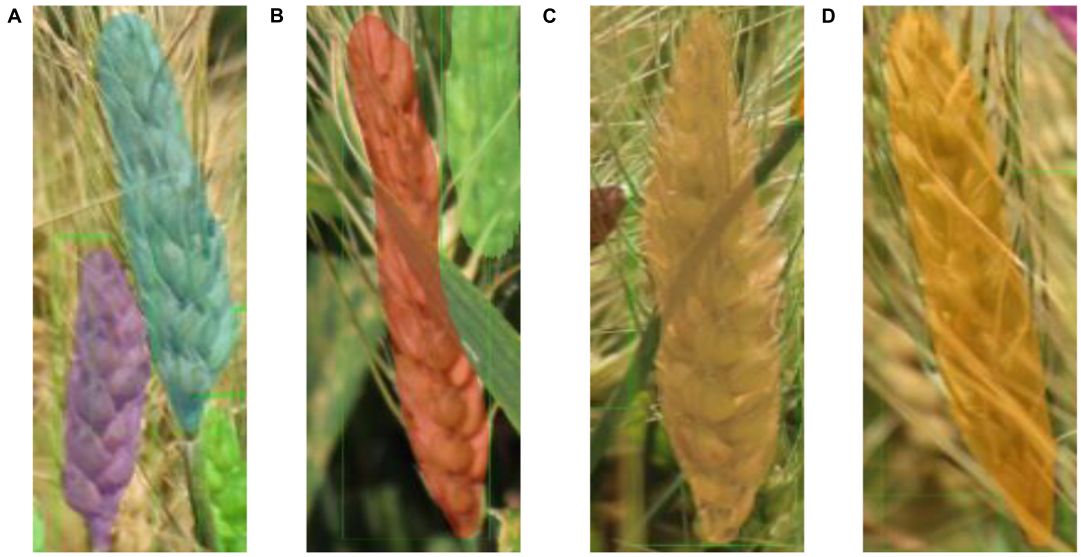


FIGURE 9 | Detection results of various occlusions. **(A)** One spike (under purple mask) is occluded by another spike, **(B)** one spike (under red mask) is occluded by a wheat leaf, **(C)** one spike is occluded by a wheat stem, **(D)** One spike (under orange mask) is occluded by awns from another spike.

TABLE 2 | The results of ablation study.

Multi-scale	DCN	GRoIE	AP (IOU = 0.5)		AP (IOU = 0.75)		Epoch	Train time/h	Test time/s
			Bbox	Mask	Bbox	Mask			
–	–	–	0.868	0.872	0.722	0.677	20	7.5	2,988
√	–	–	0.891	0.899	0.772	0.745	40	14	4,775
√	√	–	0.897	0.904	0.794	0.768	40	14	4,791
√	√	√	0.904	0.907	0.790	0.747	38	16	2,560

TABLE 3 | Comparative experimental results.

Type	AP (IOU = 0.5)		AP (IOU = 0.75)		Epoch	Train time/h	Test time/s
	Bbox	Mask	Bbox	Mask			
ResNet50	0.894	0.897	0.761	0.693	60	17	2,393
ResNet101	0.871	0.876	0.696	0.632	20	7	3,302
ResNeXt50	0.893	0.897	0.788	0.733	66	18	2,299
ResNeXt101	0.872	0.875	0.706	0.624	40	14	2,251
Res2Net50	0.904	0.907	0.790	0.747	38	16	2,560
NMS	0.904	0.907	0.790	0.747	38	16	2,560
Soft-NMS	0.903	0.906	0.795	0.750	38	16	3,385
CosineAnnealingLR	0.891	0.895	0.786	0.753	70	18	3,134
MultiStepLR	0.904	0.907	0.790	0.747	38	16	2,560
Wheat-Net	0.904	0.907	0.790	0.747	38	16	2,560
Optimized mask RCNN	0.884	0.884	0.755	0.690	60	10	3,186
Optimized cascade mask RCNN	0.899	0.900	0.785	0.754	40	14	2,673

we selected ResNet50, ResNet101, ResNeXt50, ResNeXt101, and Res2Net50 for comparative experiments. By comparing the results of ResNet50 and ResNet101 (or ResNeXt50 and ResNeXt101), we found that increasing the depth of the backbone could not improve the performance of wheat spike detection. In general, the more layers the deep neural network has, the stronger the fitting ability of the model will be. In practice, there is not only ground truth but also noise in the image. The stronger the fitting ability of the model, the stronger the ability to learn noise. In particular, the noise in this paper (such as blurry spikes, leaf, stem, and awns) is similar to the ground truth in color and texture, which makes it more difficult

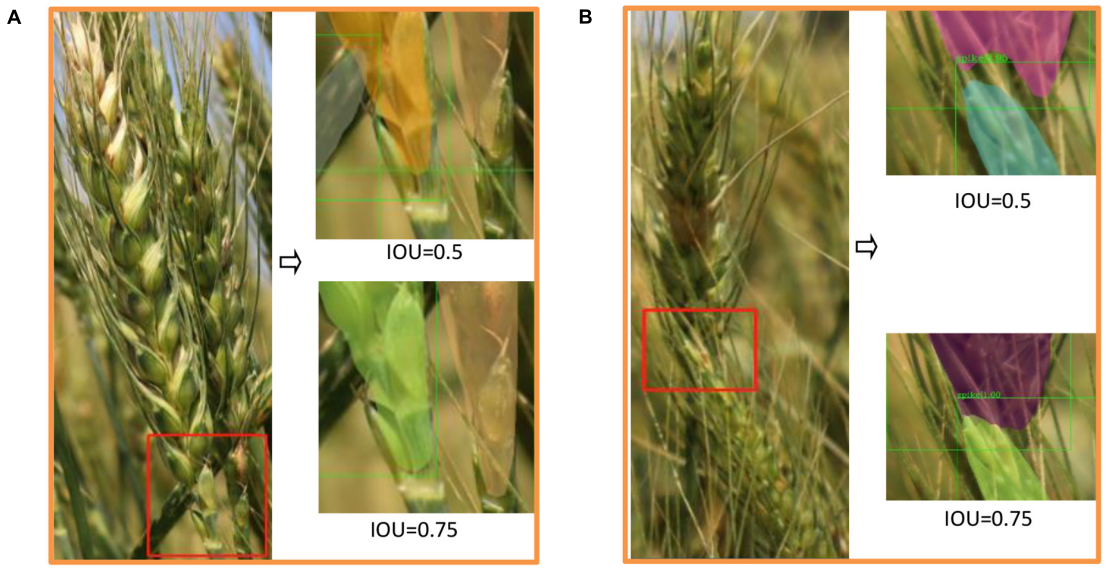


FIGURE 10 | The Demonstration of experimental error. **(A)** The error of spike bottom. **(B)** The error at the junction of different spikes.

for the model to distinguish between noise and ground truth. Therefore, in the situation of this paper, just increasing the depth of the backbone may not represent a better effect. The test results (**Table 3**) showed that Res2Net50 and ResNeXt50 have more outstanding performance than other backbone networks. Furthermore, ResNeXt50 required a shorter test time, while Res2Net50 had higher AP values. Both of the two networks can be used as the backbone of our model based on different criteria. In this paper, based on the requirement of precision, we chose Res2Net50 as the backbone network of Wheat-Net for segmenting wheat spikes in the complex field. In practical applications, if there is a higher requirement for the model speed, ResNeXt50 will be suitable to be the backbone network.

In object detection, the model will generate a lot of region proposals, and the suppression algorithm is needed to remove redundant region proposals to reduce the number of parameters in the model. In this paper, we conducted a comparative experiment between NMS and Soft-NMS to achieve the better performance of Wheat-Net. As shown in **Table 3**, the Wheat-Net with NMS achieved a higher AP with IOU = 0.5 within a much shorter test time compared to Soft-NMS. Therefore, although Soft-NMS could help the Wheat-Net to achieve slightly higher AP with IOU = 0.75, we chose to use NMS based on the best balance between precision and speed.

In the training process, if the LR is too large, the model will be difficult to converge, if the LR is too small, the convergence speed will be slow. Therefore, the dynamic decay of LR is extremely important to make the model faster and more stable to convergence. In order to choose a more suitable method of LR decay, we conducted a comparative experiment between MultiStepLR and CosineAnnealingLR to select the suitable LR decay method for Wheat-Net. From **Table 3**, we can see that the MultiStepLR was superior to CosineAnnealingLR in terms of AP. In addition, MultiStepLR converged faster and required a shorter test time than CosineAnnealingLR. Therefore, MultiStepLR was better than CosineAnnealingLR in terms of accuracy and speed for our model of wheat spike detection, so we chose MultiStepLR to decay the learning rate and further improve the performance of our model.

In order to evaluate the advantages of the hybrid cascade structure of Wheat-Net, we first used the same optimization method (including multi-scale training, DCN, and GRoIE) to optimize the Mask RCNN and Cascade Mask RCNN, and then conducted a comparative experiment. As shown in **Table 3**, we can seen that the box AP and mask AP of the Wheat-Net are better than the other two models. In addition, although the train time of the Wheat-Net was slightly longer, the Wheat-Net was more satisfactory in terms of test time and converged in the lowest number of epochs. The above analysis proves that the hybrid cascade structure of Wheat-Net is very effective for segmenting wheat spikes in the field environment.

DISCUSSION

Analysis of Experimental Error

Although Wheat-Net showed excellent performance for wheat segmentation in the complex environment, there were still errors, which we subsequently analyzed. As shown in **Figure 10A**, the

TABLE 4 | Comparison of the AP of Wheat-Net vs. other models in wheat detection.

Model	AP (IOU = 0.5)	
	Bbox	Mask
Su et al. (2021)	0.567	0.572
Wheat-Net	0.904 (↑**0.337**)	0.907 (↑**0.335**)

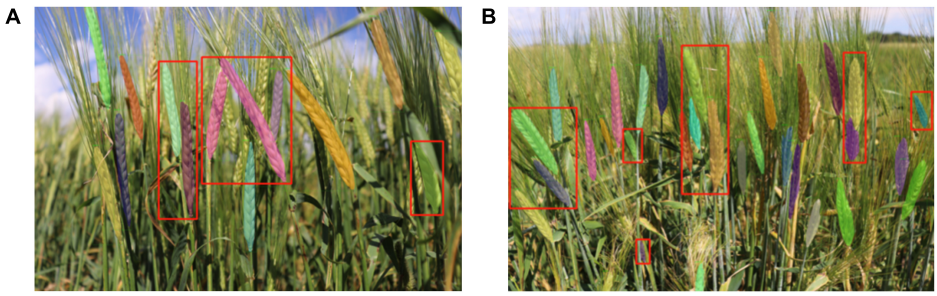


FIGURE 11 | The visualization results of barley detection. **(A,B)** The two examples of barley detection.

model had some segmentation errors at the bottom of the wheat spikes. The sparse florets at the bottom of the spike led to some differences between the texture characteristics of the bottom and other parts. This in turn caused inaccurate segmentation for the bottom of the spikes.

Due to the complexity of our dataset, the problem of adjacence and occlusion of wheat was very common in most images. Segmentation of adjacent objects was one of the most challenging tasks in the field of crop phenotyping. From **Figure 10B**, we can see that the two spikes were adhesive and the lower one was occluded by a wheat stem. Our method achieved a good segmentation result in such a complex situation, but there were still errors at the junction. The color, texture, and shape of the adherent spikes were very similar, which made the dividing line unclear. As a result, this made the positive objects at the junction annotated as negative, which increased the number of False Negatives (FN) and reduced Recall. When multi-scale training and DCN were used (GROIE was not used), the AP value with IOU of 0.75 was the highest (**Table 3**). We chose this model for visual testing (shown in **Figure 10**) and found that it could greatly reduce the above-mentioned experimental errors (including the bottom and junction errors) compared to when IOU = 0.5. Regardless of the ability to detect wheat spikes, it achieved better performance for accurately segmenting the wheat spike.

Evaluation of Wheat-Net on Barley Spike Detection

The phenotypic characteristics of wheat and barley are quite different in both the shape and size of the kernel and the length of the awn. In order to verify the generalized applicability of the model, we constructed a test set containing 29 barley images to test the detection ability of the model to barley spikes. The experimental results showed that the AP of bbox and mask for barley detection achieved 0.799 and 0.812, respectively. From **Figure 11**, we can see that the model achieved acceptable visualization results for barley, especially for the detection of adjacence and occlusion (red boxes in **Figure 11**). Thus, our model has the potential to segment barley spikes as well demonstrating strong robustness to a variety of spike shapes and colors. However, due to the similar phenotypic characteristics of adhesive spikes, there were still errors at the junction of spikes. In addition, similar to the errors encountered with wheat spike detection, there were also some errors at the top and bottom of barley spikes. It is expected that the performance of segmenting barley spikes will be improved by retraining our model using a barley training dataset. This study also established the protocol of a pretraining model for the detection of other inflorescences of small grain cereal crops such as the panicles of oat and rice.

Comparison of Wheat Detection Methods

Compared with the conventional wheat detection methods (Alkhudaydi et al., 2019; Qiu et al., 2019; Zhang et al., 2019; Ma et al., 2020; Tan et al., 2020; Su et al., 2021), the proposed Wheat-Net in this paper showed a preferable performance for instance segmentation in various complex scenes, including complex backgrounds, insufficient illumination, dense wheat spikes, spike adjacency, and occlusion. In our previous research (Su et al., 2021), we basically realized the instance segmentation of wheat in a complex field by using only a single Mask RCNN. However, since the method with a single Mask RCNN cannot learn sufficient features from our complex datasets, it had a poor effect (especially for segmentation of partial spikes of the image edge and occlusive spikes). Therefore, we spent a year devoted to improving the performance of previous study. Eventually, compare with only a single Mask RCNN of Su et al. (2021), we cascaded three Mask RCNN to construct the Wheat-Net of hybrid cascade structure, and with Res2Net50 as the backbone network, multi-scale training, DCN, and GRoIE were used to learn abundant features of different scales. From the bold values in **Table 4**, we can see that the AP of bbox is increased by 0.337 (from 0.567 to 0.904), and the AP of mask is increased by 0.335 (from 0.572 to 0.907). In addition, we can see from **Figures 8**, **9** that Wheat-Net achieved excellent performance for partial spikes of the image edge and occlusive spikes, which had a poor effect on the method of Su et al. (2021). The above experiment shows that our method can overcome various challenges in the complex field and achieve accurate and efficient instance segmentation of wheat spikes.

CONCLUSION

Due to the high complex dataset (complex backgrounds, serious occlusion), an effective instance segmentation method based on

the HTC model was established to automatically segment wheat spikes in the fields. The proposed method with a hybrid cascade structure to make full use of rich mask and box information. With Res2Net50 as the backbone network, multi-scale training was used to learn features of different scales, and deformable convolutional networks (DCN) and Generic RoI Extractor (GRoIE) were trained to improve model accuracy. Based on the methodology, the difficulties of complex backgrounds, serious occlusion, and incomplete spikes on the edge were solved with AP of 0.904 and 0.907 for bbox and mask, respectively. The accuracy rate for wheat spike counting was 99.29%. Comprehensive empirical analyses revealed that the proposed method was particularly effective for the detection of wheat spikes with frequent adjacence, overlapping, occlusion, and other complex growth states. This study achieved excellent performance for dense wheat spike segmentation with complex field, which is conducive to promoting production and management of wheat. However, field data collection is limited to only the crop season, which is 3 months per year in Minnesota. One solution is to expand the data collection window by conducting multi-site data collection across regions. In addition, we will study the method of data augmentation based on the Generative Adversarial Network (GAN). Our models will also be used in the large-scale wheat field trials. We expect that our proposed method will be expanded to the broader agricultural research area, including detection of the seed-bearing inflorescences of other crops.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JZ, AM, CY, and BS: conceptualization. JZ, AM, QM, and CY: methodology and formal analysis. JZ, AM, CY, and JW: software. JZ, QM, AM, CY, and JW: validation. CY, CH, W-HS, and BS: investigation. JZ, AM, CY, CH, JA, and BS: data curation. JZ: writing-original draft preparation. JZ, AM, CY, QM, BS, CH, and JA: writing-review and editing. CY, BS, W-HS, and QM: project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the USDA-ARS United States Wheat and Barley Scab Initiative (Grant No. 59-0206-0-181), the Lieberman-Okinow Endowment at the University of Minnesota, and the State of Minnesota Small Grains Initiative. The research was also supported by Provincial Natural Science Foundation Project (Grant No. ZR2021MC099).

REFERENCES

Alkhudaydi, T., Reynolds, D., Griffiths, S., Zhou, J., and Iglesia, B. (2019). An exploration of deep-learning based phenotypic analysis to detect spike regions in field conditions for UK bread wheat. *Plant Phenomics* 2019, 1–17. doi: 10.34133/2019/7368761

Aparicio, N., Villegas, D., Casadesus, J., Araus, J. L., and Royo, C. (2000). Spectral vegetation indices as nondestructive tools for determining durum wheat yield. *Agron. J.* 92, 83–91. doi: 10.2134/agronj2000.92183x

Bodla, N., Singh, B., Chellappa, R., and Davis, L. S. (2017). "Soft-NMS — improving object detection with one line of code," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, (Venice: IEEE), 5562–5570.

Cai, Z., and Vasconcelos, N. (2021). Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intellig.* 43, 1483–1498. doi: 10.1109/TPAMI.2019.2956516

Caruana, R. (1997). Multitask learning. *Mach. Learn.* 28, 41–75. doi: 10.1023/A:1007379606734

Chandra, A., Desai, S. V., Balasubramanian, V., Ninomiya, S., and Guo, W. (2020). Active learning with point supervision for cost-effective panicle detection in cereal crops. *Plant Methods* 16:34. doi: 10.1186/s13007-020-00575-8

Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., et al. (2019). "Hybrid task cascade for instance segmentation," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: IEEE), 4969–4978.

Cointault, F., Guerin, D., Guillemin, J., and Chopinet, B. (2008a). In-field Triticum aestivum ear counting using colour-texture image analysis. *N. Zeal. J. Crop Hortic. Sci.* 36, 117–130. doi: 10.1080/01140670809510227

Cointault, F., Journaux, L., Destain, M.-F., and Gouton, P. (2008b). "Wheat ear detection by textural analysis for improving the manual countings," in *Proceedings of the 5th IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA)*, Innsbruck.

da Silva, L. A., Bressan, P. O., Gonçalves, D. N., Freitas, D. M., Machado, B. B., and Gonçalves, W. N. (2019). Estimating soybean leaf defoliation using convolutional neural networks and synthetic images. *Comput. Electron. Agric.* 156, 360–368. doi: 10.1016/j.compag.2018.11.040

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., et al. (2017). "Deformable convolutional networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, (Piscataway, NJ: IEEE), 764–773.

David, E., Madec, S., Sadeghi-Tehran, P., Aasen, H., Zheng, B., Liu, S., et al. (2020). Global Wheat Head Detection (GWHD) dataset: a large and diverse dataset of high resolution RGB labelled images to develop and benchmark wheat head detection methods. *Plant Phenomics* 2020:3521852. doi: 10.34133/2020/3521852

David, E., Serouart, M., Smith, D., Madec, S., Velumani, K., Liu, S., et al. (2021). Global Wheat Head Dataset 2021: More diversity to improve the benchmarking of wheat head localization methods. *Plant Phenomics* 2021:9846158.

de Luna, R., Dadios, E., Bandala, A., and Vicerra, R. (2020). Tomato growth stage monitoring for smart farm using deep transfer learning with machine learning-based maturity grading. *Agrivita J. Agric. Sci.* 42, 24–36. doi: 10.17503/agrivita.v42i1.2499

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 the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, (Piscataway, NJ: IEEE), 248–255. doi: 10.1109/TMI.2016.2528162

Desai, S. V., Balasubramanian, V. N., Fukatsu, T., Ninomiya, S., and Guo, W. (2019). Automatic estimation of heading date of paddy rice using deep learning. *Plant Methods* 15:76. doi: 10.1186/s13007-019-0457-1

Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. (2021). Res2Net: a new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intellig.* 43, 652–662. doi: 10.1109/TPAMI.2019.2938758

Germain, C., Rousseaud, R., and Grenier, G. (1995). "Non destructive counting of wheatear with picture analysis," in *Proceedings of the Fifth International Conference on Image Processing and Its Applications*, (Piscataway, NJ: IEEE), 435–439.

Hasan, M. M., Chopin, J. P., Laga, H., and Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* 14:100. doi: 10.1186/s13007-018-0366-8

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," in *Procedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, (Piscataway, NJ: IEEE), 2980–2988.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intellig.* 37, 1904–1916. doi: 10.1109/TPAMI.2015.2389824

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: IEEE), 770–778.

Khaki, S., Wang, L., and Archontoulis, S. (2020). A CNN-RNN framework for crop yield prediction. *Front. Plant Sci.* 10:1750. doi: 10.3389/fpls.2019.01750

Kovalchuk, N., Laga, H., Cai, J., Kumar, P., Parent, B., Lu, Z., et al. (2017). Phenotyping of plants in competitive but controlled environments: a study of drought response in transgenic wheat. *Funct. Plant Biol.* 44, 290–301. doi: 10.1071/FP16202

Li, X., Liu, Z., Luo, P., Loy, C. C., and Tang, X. (2017). "Not all pixels are equal: difficulty-aware semantic segmentation via deep layer cascade," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: IEEE), 6459–6468.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft COCO: common objects in context," in *Computer Vision – ECCV 2014*, eds D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Berlin: Springer International Publishing), 740–755. doi: 10.1089/big.2021.0262

Ma, J., Li, Y., Liu, H., Du, K., Zheng, F., Wu, Y., et al. (2020). Improving segmentation accuracy for ears of winter wheat at flowering stage by semantic segmentation. *Comput. Electron. Agric.* 176:105662. doi: 10.1016/j.compag.2020.105662

Misra, T., Arora, A., Marwaha, S., Chinnusamy, V., Rao, A., Jain, R., et al. (2020). SpikeSegNet-a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* 16:40. doi: 10.1186/s13007-020-00582-9

Ni, C., Wang, D., Vinson, R., Holmes, M., and Tao, Y. (2019). Automatic inspection machine for maize kernels based on deep convolutional neural networks. *Biosyst. Eng.* 178, 131–144. doi: 10.1016/j.biosystemseng.2018.11.010

Pont-Tuset, J., Arbeláez, P., Barron, T., Marques, F., and Malik, J. (2017). Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Trans. Pattern Anal. Mach. Intellig.* 39, 128–140. doi: 10.1109/TPAMI.2016.2537320

Pound, M., Atkinson, J., Wells, D., Pridmore, T., and French, A. (2017). Deep learning for multi-task plant phenotyping. *bioRxiv* [Preprint]. doi: 10.1101/204552

Qiongyan, L., Cai, J., Berger, B., Okamoto, M., and Miklavcic, S. J. (2017). Detecting spikes of wheat plants using neural networks with Laws texture energy. *Plant Methods* 13:83. doi: 10.1186/s13007-017-0231-1

Qiu, R., Yang, C., Moghimi, A., Zhang, M., Steffenson, B., and Hirsch, C. (2019). Detection of fusarium head blight in wheat using a deep neural network and color imaging. *Remote Sens.* 11:2658. doi: 10.3390/rs11222658

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. Intellig.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Rosenfeld, A., and Thurston, M. (1971). Edge and curve detection for visual scene analysis. *IEEE Trans. Comput.* 20, 562–569. doi: 10.1109/T-C.1971.223290

Rossi, L., Karimi, A., and Prati, A. (2020). "A novel region of interest extraction layer for instance segmentation," in *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)*, (Piscataway, NJ: IEEE).

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. Vis.* 77, 157–173. doi: 10.1007/s11263-007-0090-8

Sadeghi-Tehran, P., Virlet, N., Ampe, E., Reyns, P., and Hawkesford, M. (2019). DeepCount: in-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks. *Front. Plant Sci.* 10:1176. doi: 10.3389/fpls.2019.01176

Senthilkumar, T., Jayas, D. S., White, N. D. G., Fields, P. G., and Gräfenhan, T. (2017). Detection of ochratoxin A contamination in stored wheat using near-infrared hyperspectral imaging. *Infrared Phys. Technol.* 81, 228–235. doi: 10.1016/j.infrared.2017.01.015

Shen, Y., Yin, Y., Li, B., Zhao, C., and Li, G. (2021). Detection of impurities in wheat using terahertz spectral imaging and convolutional neural networks. *Comput. Electron. Agric.* 181:105931. doi: 10.1016/j.compag.2020.105931

Su, W.-H., Zhang, J., Yang, C., Page, R., Szinyei, T., Hirsch, C., et al. (2021). Automatic evaluation of wheat resistance to fusarium head blight using dual mask-RCNN deep learning frameworks in computer vision. *Remote Sens.* 13, 1–21. doi: 10.3390/rs13010026

Sun, J., Di, L., Sun, Z., Shen, Y., and Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors* 19:E4363. doi: 10.3390/s19204363

Tan, C., Zhang, P., Zhang, Y., Zhou, X., Wang, Z., Ying, D., 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

Wang, Y., Qin, Y., and Cui, J. (2021). Occlusion robust wheat ear counting algorithm based on deep learning. *Front. Plant Sci.* 12:645899. doi: 10.3389/fpls.2021.645899

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). "Aggregated residual transformations for deep neural networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Piscataway, NJ: IEEE), 5987–5995.

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., et al. (2020). Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant Methods* 16:106. doi: 10.1186/s13007-020-00648-8

Yang, Q., Shi, L., and Lin, L. (2019). "Plot-scale rice grain yield estimation using UAV-based remotely sensed images via CNN with time-invariant deep features decomposition," in *Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, (Piscataway, NJ: IEEE), 7180–7183.

Zhang, D., Wang, D., Gu, C., Jin, N., Zhao, H., Chen, G., et al. (2019). Using neural network to identify the severity of wheat fusarium head blight in the field environment. *Remote Sens.* 11:2375. doi: 10.3390/rs11202375

Zhang, J., Ma, Q., Cui, X., Guo, H., Wang, K., and Zhu, D. (2020). High-throughput corn ear screening method based on two-pathway convolutional neural network. *Comput. Electron. Agric.* 175:105525. doi: 10.1016/j.compag.2020.105525

Zhuang, S., Wang, P., Jiang, B., and Li, M. (2020). Learned features of leaf phenotype to monitor maize water status in the fields. *Comput. Electron. Agric.* 172:105347. doi: 10.1016/j.compag.2020.105347

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fpls.2022.806878

YOLO-VOLO-LS: A Novel Method for Variety Identification of Early Lettuce Seedlings

Pan Zhang[1,2,3,4,5] *and Daoliang Li*[1,2,3,4,5]*

[1] National Innovation Center for Digital Fishery, China Agricultural University, Beijing, China, [2] Beijing Engineering and Technology Research Centre for Internet of Things in Agriculture, China Agriculture University, Beijing, China, [3] China-EU Center for Information and Communication Technologies in Agriculture, China Agriculture University, Beijing, China, [4] Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, China Agriculture University, Beijing, China, [5] College of Information and Electrical Engineering, China Agricultural University, Beijing, China

Accurate identification of crop varieties is an important aspect of smart agriculture, which is not only essential for the management of later crop differences, but also has a significant effect on unmanned operations in planting scenarios such as facility greenhouses. In this study, five kinds of lettuce under the cultivation conditions of greenhouses were used as the research object, and a classification model of lettuce varieties with multiple growth stages was established. First of all, we used the-state-of-the-art method VOLO-D1 to establish a variety classification model for the 7 growth stages of the entire growth process. The results found that the performance of the lettuce variety classification model in the SP stage needs to be improved, but the classification effect of the model at other stages is close to 100%; Secondly, based on the challenges of the SP stage dataset, we combined the advantages of the target detection mechanism and the target classification mechanism, innovatively proposed a new method of variety identification for the SP stage, called YOLO-VOLO-LS. Finally, we used this method to model and analyze the classification of lettuce varieties in the SP stage. The result shows that the method can achieve excellent results of 95.961, 93.452, 96.059, 96.014, 96.039 in Val-acc, Test-acc, Recall, Precision, F1-score, respectively. Therefore, the method proposed in this study has a certain reference value for the accurate identification of varieties in the early growth stage of crops.

Keywords: hydroponic crops, greenhouse, deep learning, detection, classification, multiple growth stages

OPEN ACCESS

Edited by:
Zhao Zhang,
North Dakota State University,
United States

Reviewed by:
Saeed Hamood Alsamhi,
Ibb University, Yemen
Mohsen Yoosefzadeh Najafabadi,
University of Guelph, Canada

***Correspondence:**
Daoliang Li
dliangl@cau.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *01 November 2021*
Accepted: *04 February 2022*
Published: *24 February 2022*

Citation:
Zhang P and Li D (2022) YOLO-VOLO-LS: A Novel Method for Variety Identification of Early Lettuce Seedlings. Front. Plant Sci. 13:806878. doi: 10.3389/fpls.2022.806878

INTRODUCTION

With the integration of modern information technology such as artificial intelligence, big data, and the Internet of Things with agricultural development, smart agriculture has become the inevitable direction of agricultural development (Kussul et al., 2017; Mei et al., 2018). As one of the important contents of the development of smart agriculture, the intelligent identification and classification of crop varieties is crucial to the management of the differences in later crop production (Suh et al., 2018; Khamparia et al., 2020). In addition, there are certain differences between different varieties of the same crop in terms of growth cycle, fertilizer requirements, light requirements, heat resistance, cold resistance, etc. (Yalcin and Razavi, 2016). If they are not distinguished and identified, they will

face many problems in later production management. The traditional crop variety identification process mostly relies on human identification by experts and planters, which is time-consuming, laborious and inefficient (Sun et al., 2017), and it is difficult to automatically connect tasks such as intelligent irrigation, fertilization, grading, sorting, packaging, and harvesting in the planting process. Therefore, there is an urgent need for an intelligent identification method to automate the identification of crop varieties in order to realize the unmanned connection of related tasks in the whole production process.

The advent of artificial intelligence provides a strong technical support for the intelligent identification of crop varieties (Tan et al., 2020). Image processing, machine learning, and deep learning have been continuously integrated and applied with agricultural research (Kim et al., 2018; Kaya et al., 2019). As far as crop varieties recognition is concerned, traditional image processing methods have been applied earlier (Dharwadkar et al., 2017; Tiwari, 2020). However, traditional image processing methods require researchers to manually design and extract features, such as the color, shape, and texture of crop leaves. There is a certain degree of blindness in this process (Bhosle and Musande, 2019). At the same time, based on the manually extracted feature data, neural networks composed of neurons, such as multi-layer perceptron, are used for model construction (Zhang et al., 2020). Many parameters need to be manually adjusted in the modeling process, and the model is easy to over fit, which increases the difficulty of model construction (Yoosefzadeh-Najafabadi et al., 2021). With the rise of artificial intelligence algorithms such as deep learning, it has gradually made breakthroughs in progress in the fields of computer vision, image classification, target detection, target segmentation, and speech recognition (Yalcin and Razavi, 2016). Deep learning is a new field of machine learning, which automatically analyzes data and extracts features by simulating the brain (Tóth et al., 2016; Khamparia et al., 2020). In data processing, a neural network for the target task is established through the basic CNN network, and the characteristics of the input data (such as color, texture, shape, etc.) are extracted layer by layer, and a good mapping relationship from the underlying signal to the high-level semantics is established (Dileep and Pournami, 2019). Therefore, deep learning may have more advantages in crop varieties identification.

In recent years, research on the recognition of crop varieties or types based on deep learning methods are mostly in field planting scenarios (Grinblat et al., 2016; Teimouri et al., 2019). On the one hand, for agronomists and agricultural institutions specializing in land management, it is very important to fully understand the specific conditions of land use and dynamically monitor crop planting within a certain period of time. For example, Mazzia et al. (2020) used remote sensing technology to obtain multi-temporal sentinel-2 images in central and northern Italy and combined recurrent neural network (RNN) and convolutional neural network (CNN) to propose a pixel-based LC&CC deep learning model for the region's type identification of agricultural crops. By comparing traditional support vector machine, random forest, and other methods, the accuracy of the proposed LC&CC deep learning method can reach 96.5%. In order to achieve accurate, automatic, and rapid crop mapping, Sun et al. (2020) built a deep neural network classification model based on historical crop maps and ground measurement data in North Dakota, and a high-quality map of seasonal crops was generated from Landsat images of North Dakota. At the same time, when the model was applied to new images, accurate results were obtained on major crops such as corn, soybeans, barley, spring wheat, dried beans, sugar beets, and alfalfa. For the problem of spectral similarity between different plants in the same family and genus, Zhang et al. (2020) used an improved three-dimensional CNN to build a tree species classification model based on a remote sensing data set with rich spectral and spatial characteristics. The results show that this method can reach 93.14% accuracy. In order to better capture the temporal and spatial characteristics of crop classification, Gadiraju et al. (2020) proposed a multi-modal deep learning method that combines spatial spectrum and phenological characteristics. Among them, the spatial characteristics of the image are obtained through CNN, and the phenological characteristics of the image are obtained through LSTM. The results show that this method can reduce the error by 60%. In addition, the accurate identification of agricultural products varieties is not only an urgent need of dealers, but also an urgent need of product processing enterprises and consumers. Rong et al. (2020) collected the visible and near-infrared spectrum data of five peach varieties between 350 and 820 nm, and then constructed a one-dimensional CNN to identify peach varieties with an accuracy of 94.4%. Liu et al. (2020) used machine learning and computer vision technology to classify 7 kinds of chrysanthemum tea. Compared with traditional morphological feature extraction (90%), the classification performance of deep neural network is better (96%). Liu et al. (2019) used VGG16 and ResNet50 to identify chrysanthemum varieties, which further proved that the deep learning method applied to variety recognition research has the advantages of strong recognition performance and fast recognition speed, which is a breakthrough in horticultural science. Van Hieu and Hien (2020) obtained images of 109 Vietnamese plants through the Vietnam Encyclopedia, and then used MobileNetV2, VGG16 and other methods to construct classification models. The results showed that MobilenetV2 has the highest recognition rate of 83.9%. Bisen (2021) built a recognition and classification system for different crops based on leaves, and extracted leaf features through a CNN, and finally got an accuracy of 93.75%. Nkemelu et al. (2018) compared the classification performance of two traditional methods and CNNs based on image data sets of 960 plant species at 12 different growth stages. The results show that reasonable use of CNNs can achieve ideal classification results. Grinblat et al. (2016) used a deep CNN to build a crop classification model based on the leaf vein patterns of three bean crops, and the results showed that the effect of the leaf vein-based crop classification model has been significantly improved. At the same time, it has been proved that increasing the network depth can further improve the effect of the model. Tan et al. (2020) also constructed a crop model based on the vein characteristics of plants, and achieved good results. At the same time, the effectiveness of leaf vein characteristics in the

process of plant classification was also proved by Lee et al. (2017). In large-scale plant species identification and classification, in order to improve the accuracy and computational efficiency of plant species identification, Zhang H. et al. (2018) proposed a path-based tree classifier deep learning method. The classification is carried out in a detailed hierarchical structure, and the effect is significantly improved. Similarly, some researchers have used MaskRCNN, AlexNet, CNN and other methods to identify each varieties of bananas (Le et al., 2019), grapes (Pereira et al., 2019), lemons (Alzamily et al., 2019) and medicinal materials (Dileep and Pournami, 2019; Duong-Trung et al., 2019), and achieved good results. On the other hand, in some countries where small farms are the main planting model, there are more small-scale land, dense intercropping, and diverse crop types. Chew et al. (2020) obtained image data through drones and built a recognition model for bananas, corn, beans and other crops based on VGG, and achieved good results. However, in the case of crop intercropping, there are certain limitations in the recognition accuracy of different crops. Synthetic aperture radar data also has certain advantages in remote sensing crop recognition. Teimouri et al. (2019) proposed a new method—FCN-LSTM by combining full convolutional network (FCN) and long short-term memory network (LSTM). This method has been applied to radar data to construct a remote sensing crop classification model, and the results show that the accuracy of the method in the classification of 8 crops based on pixel recognition exceeds 86%.

Based on the results of the above research, although there have been studies on species or type identification for some crops, most of the application scenarios are field planting, and a small number of application scenarios are gardening, and there are few scenarios such as greenhouses (Zhang H. et al., 2018; Teimouri et al., 2019; Gadiraju et al., 2020). The devices used mainly include spectroscopy and digital cameras (Zhang et al., 2020; Bisen, 2021). Among them, spectroscopy equipment is expensive, and it is mostly used in large-scale planting scenarios (Zhang et al., 2020). Digital cameras are relatively cheap, and can meet the needs of low-altitude remote sensing and greenhouse scenes (Bisen, 2021). However, due to the limitations of the greenhouse space and structure, there are certain risks in carrying the camera on the drone equipment, but carrying the camera on the mobile robot can achieve most of the greenhouse agricultural production tasks (Zhang et al., 2019). In addition, most of the above studies directly use deep learning classification methods (such as AlexNet, VGG, ResNet, etc.) to identify the varieties or types of different crops (Ghazi et al., 2016). In the greenhouse scenario, the initial growth of most crops is relatively small and the background features account for a large proportion. In this case, the direct classification of varieties or types of crops may cause a certain loss of accuracy. However, by combining target detection and classification, there may be unexpected results.

Therefore, this study conducted the identification of lettuce varieties at different growth stages for 5 kinds of greenhouse hydroponic lettuce under 6 nitrogen treatments. The classification method and target detection method are used for crop variety recognition, and the two are combined for lettuce variety recognition to explore the improvement of model performance. The novel contributions of this article are concluded as follows: (1) We constructed 7 lettuce classification datasets under different growth stages and different nitrogen treatments, and used the dataset to study the classification of lettuce varieties under the influence of multiple factors. (2) We used classification-based and detection-based methods for lettuce variety recognition, and compared the performance of lettuce variety recognition models at different growth stages. (3) We propose a lettuce variety recognition method called YOLO-VOLO-LS, which combines classification mechanism and detection mechanism, and discuss its challenges and opportunities in future application.

MATERIALS AND METHODS

Experimental Field

This experiment was conducted in the glass greenhouse of the Factory Agricultural Research and Development Center of Chongqing Academy of Agricultural Sciences from March to May of 2021 (**Figure 1**). Five varieties of lettuce were selected for the experiment, namely, Selected Italian (V1), Small cream green (V2), Rosa green (V3), Badawiya (V4), and Boston cream (V5). In the seedling stage, we selected the full-grained lettuce seeds and placed them in the seedling cotton with 100 grooves for seeding, with one seed in each groove. The temperature was controlled at 23–26°C, the air humidity was 60–70%, the seedling cotton was kept moist, and the halogen lamp was used to supplement the light after germination. After 20 days of seedlings, we transplanted 5 varieties of lettuce to 6 cyclically rotating stereoscopic cultivation racks. Each cultivation rack was set to 0, 33, 66, 99, 132, and 165% according to the nitrogen concentration in the standard nutrient solution. Five slots were designated on both sides of each cultivating rack to cultivate a specific variety of lettuce. Each slot can grow 81 lettuces, and the two sides of the stereoscopic cultivating rack are correspondingly placed with the same variety of lettuce (**Figure 2**). The growth process of lettuce adopted the way of hanging roots, and the nutrient solution was changed every 3 days. Normal greenhouse cultivation and management of lettuce were conducted, and no pesticides and hormones were applied.

Image Data Acquisition

The whole process from seedling to transplanting of lettuce in this experiment mainly included three stages: seedling (SL), separate planting (SP), and transplanting (TP), as shown in **Figure 3**.

First, after we separated the five varieties of lettuce from the seeding cotton, we collected data every other day for the next 12 days. We conducted collections six times in total, and 160 images of each lettuce are collected each time. Second, after the five varieties of lettuce seedlings were transplanted to the stereo cultivation rack, we carried out image data acquisition every 5 days. Each data collection mainly acquired 50 lettuce images of different varieties and different nitrogen nutrient gradient treatments, each of which had six nitrogen treatments, for a total of 300 images. The data collection time was 9:00–17:00, and in order to ensure the consistency of data collection, the camera

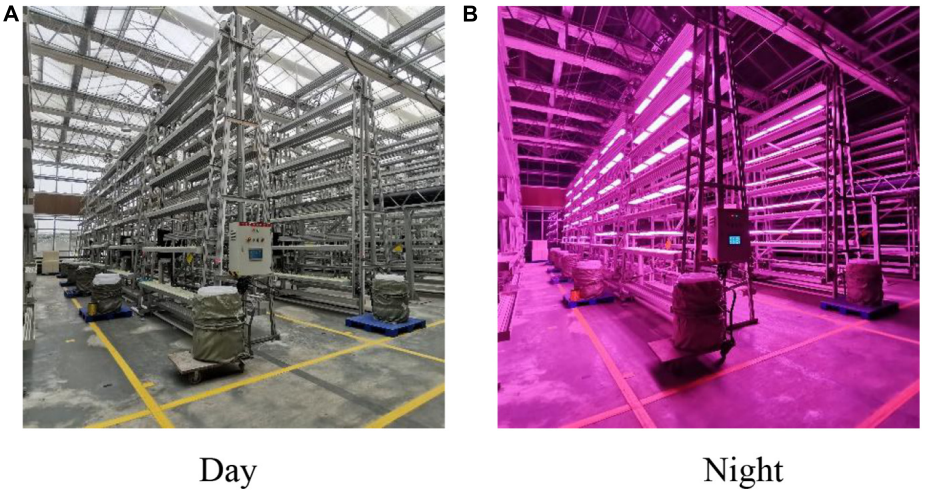


FIGURE 1 | Greenhouse cultivation environment. **(A)** Represents the cultivation environment under daytime conditions, and **(B)** represents the cultivation environment under night conditions.

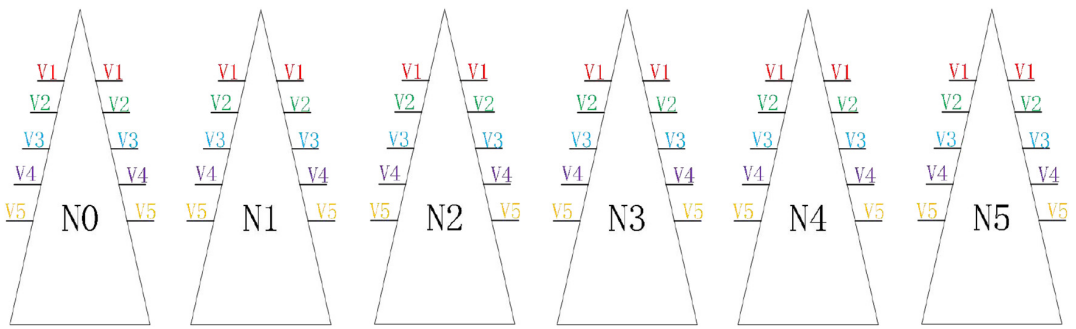


FIGURE 2 | Distribution of lettuce in stereoscopic cultivation racks of 5 varieties. There were a total of 6 cultivation racks, and each cultivation rack was equipped with a nitrogen concentration treatment, and the same variety of lettuce was symmetrically transplanted at the same position on both sides of each cultivation rack.

was kept perpendicular to the plane of the planting slots and at a distance of 40 cm during the collection process.

Data Pre-processing

Based on the aforementioned data acquisition process, the number of data acquired on Days 1, 6, 12, 18, 24, and 30 was only 300, which cannot meet the data volume requirements for deep learning training. Based on the principle of cross-validation, we first randomly divided the lettuce data set of each variety according to the ratio of 6:2:2. Second, we performed data enhancement through rotation, flipping, and contrast adjustment, 23 times. Finally, the dataset volume of the training, validation and test for each growth stage of each variety after data enhancement was 4,140, 1,380, and 1,380, respectively.

In addition, in order to ensure that our model can achieve accurate classification of lettuce varieties in various greenhouse scenarios, we used a contrast adjustment method in data enhancement to improve the richness of data. Specifically, this is as shown in **Figure 4**.

YOLO-VOLO

In this study, the recognition of the five lettuce varieties in **SP** had problems of strong background interference, high similarity, and difficulty in classification (**Figure 5**). We propose a new method called YOLO-VOLO to identify lettuce varieties in the SP stage by combining target detection and target classification mechanisms (**Figure 6**), so as to achieve a relatively ideal classification effect.

YOLO-v5 for Target Cutting

First, we use the LabelImg software to set the label for the five lettuce datasets of **SP** as Plant. Then, YOLO-v5 (Jia et al., 2021; Kasper-eulaers et al., 2021; Liu W. et al., 2021), the most advanced algorithm of Yolo series, was used to detect the lettuce plant in **SP**, namely separating from seedling cotton. After training an object detection model independently in **SP**, we cut the lettuce plants according to the coordinates of the location of each plant predicted by the target detection model. Due to the differences in the growth of each lettuce plant, the cropped objects are of different sizes. In order to ensure that each object

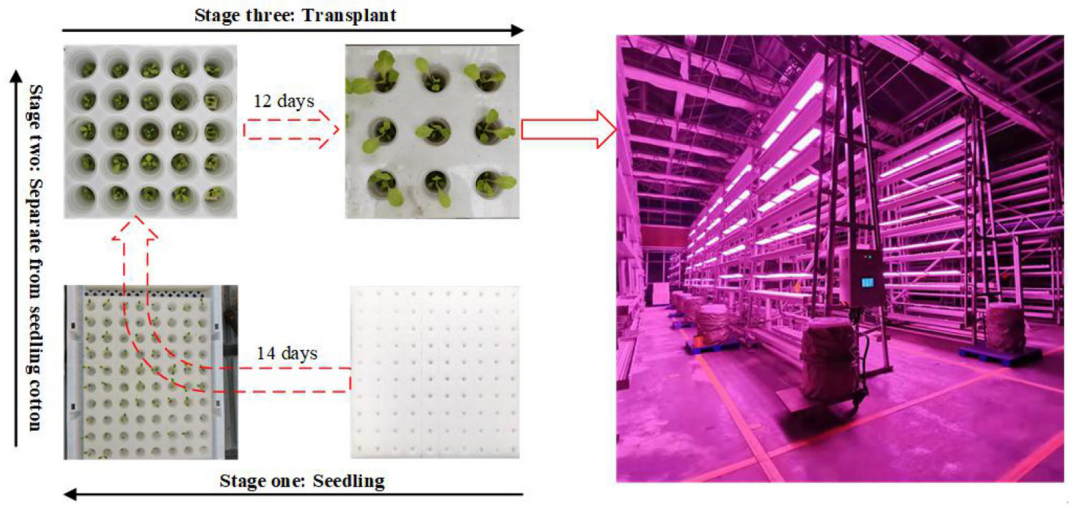


FIGURE 3 | The whole process from seedling to transplanting of the lettuce. The whole process mainly included 3 stages, namely seedling, separating from seedling cotton, and transplanting.

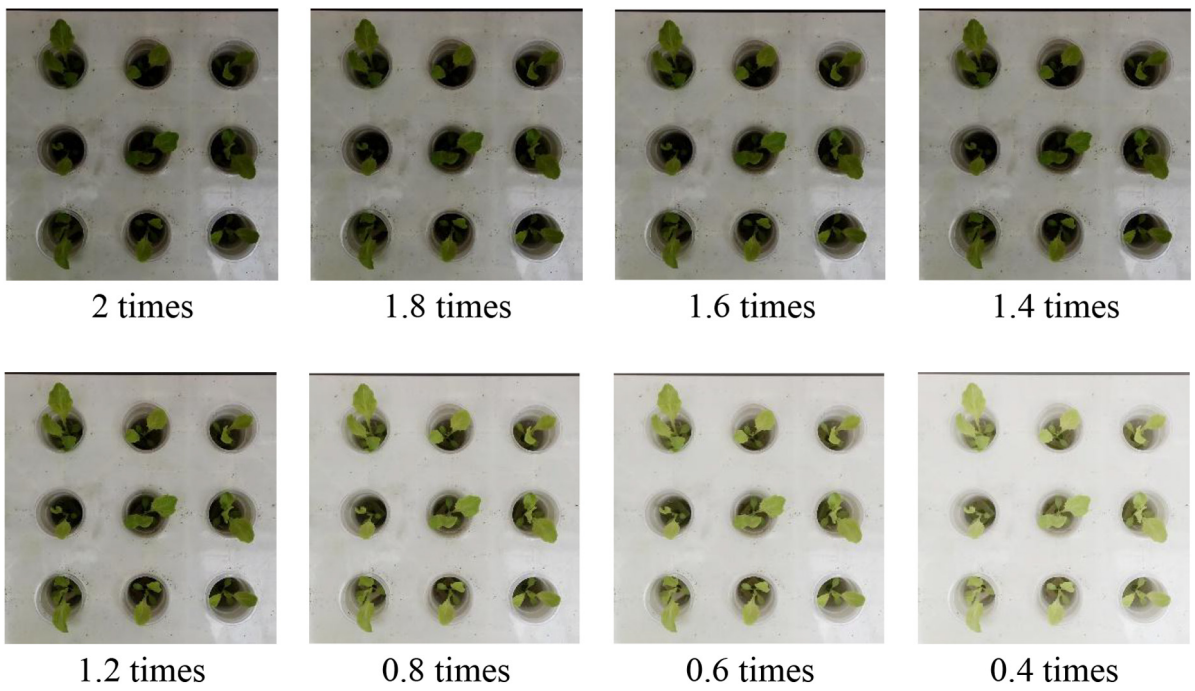


FIGURE 4 | Data enhancement based on contrast adjustment. The adjustment of the image contrast is realized by the gamma adjustment method. When gamma $>$ 1, the new image is darker than the original image. If gamma $<$ 1, the new image is brighter than the original image.

maintains the original image aspect ratio during the later model training, we used the boundary padding method to perform data preprocessing (**Figure 6A**), according to the characteristics of the image datasets of **SP** of lettuce, and considering the requirements of image resolution, GPU memory and accurate detection. We put the image datasets of the Stage two into the neural network for training (the image resolution is 384 * 384), where three different sizes of detection head, 52*52, 26*26, and 13*13 are used to output the results including the lettuce's position information, category information and confidence.

VOLO for Target Classification

Based on the lettuce images obtained from the processing in Section "YOLO-v5 for target cutting," we used the current state-of-the-art target recognition algorithm VOLO to classify the five lettuce images. VOLO is a network structure with two

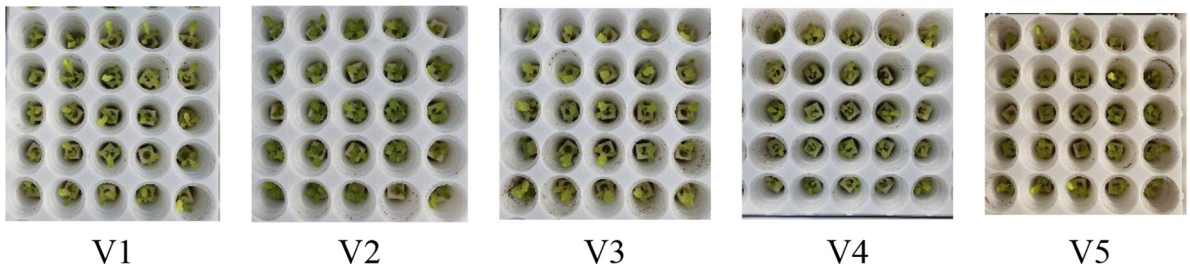


FIGURE 5 | Five kinds of lettuce involved in this study. Among them, Italian (V1), Small cream green (V2), Rosa green (V3), Badawiya (V4), and Boston cream (V5).

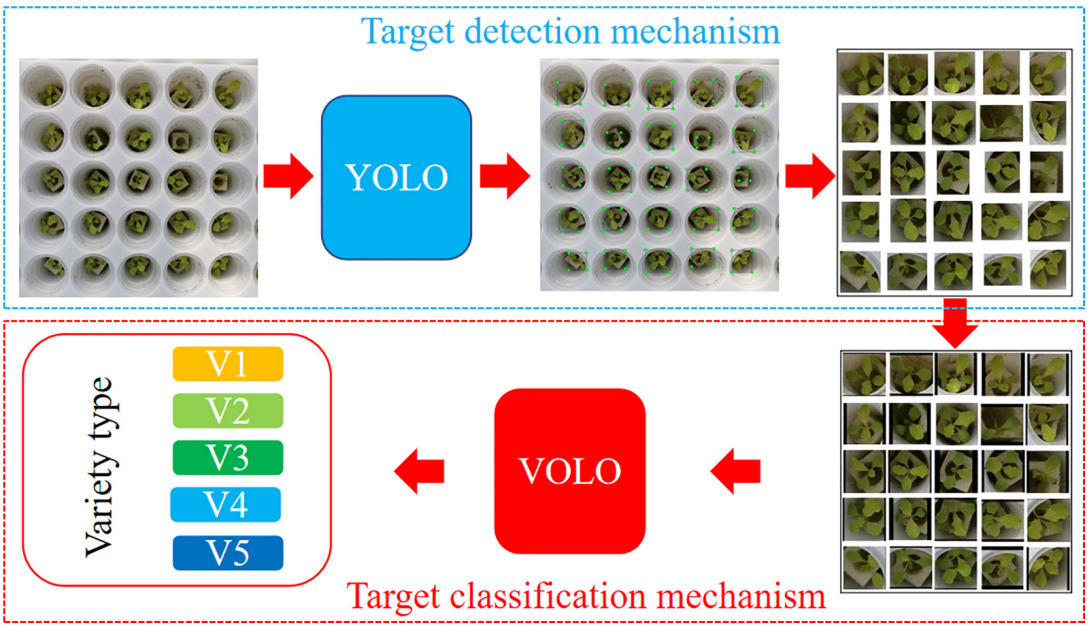


FIGURE 6 | Data processing process of the YOLO-VOLO model. The input RGB image is first detected and cropped by the YOLO algorithm, and then input into the VOLO algorithm for lettuce species identification after boundary padding, and finally the category of the lettuce species is output.

independent stage (Yuan et al., 2021). First, a stack of Outlookers that generates a fine-grained token representation constitutes the first independent stage. Secondly, a second independent stage is formed to aggregate global information by deploying a series of transformer blocks. At the beginning of each stage, a patch embedding module is used to map the input to the marked representation of the design shape.

Outlooker is a newest simple and lightweight attention mechanism module, which can effectively use fine-level information to enrich token representation. In addition, Outlooker has made certain innovations in generating attention for token aggregation, allowing the model to efficiently encode fine-level information. In particular, an effective linear mapping method can directly infer the mechanism of gathering surrounding tokens from the characteristics of anchored tokens, thereby avoiding expensive point product attention calculations. The Outlooker is composed of the outlook attention layer used for spatial information encoding and the multi-layer perceptron (MLP) used for information interaction between channels. Given a sequence of input $C-$dim token representations $X \in R^{H\times W\times C}$, Outlooker can be represent as follows:

$$\widetilde{X} = OutlookAtt(LN(X)) + X \tag{1}$$

$$Z = MLP(LN(\widetilde{X})) + \widetilde{X} \tag{2}$$

Where LN refer to LayerNorm (Liu F. et al., 2021).

Among them, Outlook attention is efficient, easy, and simple to implement. The main characteristics are: (1) The features at each spatial location are sufficiently representative to generate attention weights for local aggregation of neighboring features; (2) Dense local spatial aggregation can effectively encode fine-level information.

As we can see from **Figure 7**, for a partial window with a size of K × K, a linear layer can be simply generated from the central token, and a reshaping operation (highlighted by the green box) can then be performed. Since the attention weight is generated

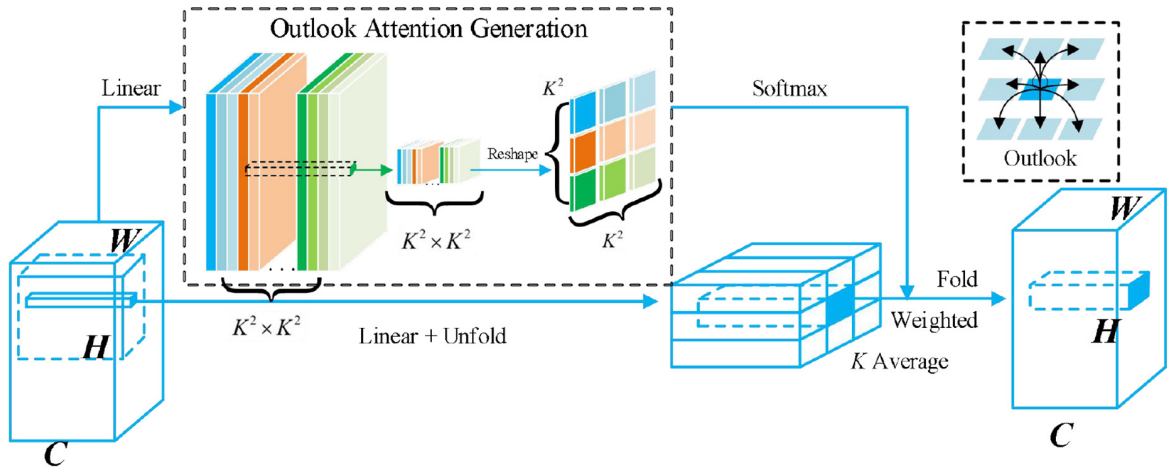


FIGURE 7 | Illustration of outlook attention. The outlook attention matrix for a local window of size K*K can be simply generated from the center token with a linear layer followed by a reshape operation. The attention weights are generated from the center token within the window and act on the neighbor tokens and itself.

by the center mark in the window and acts on the adjacent mark and itself (highlighted by the black box), we call these operations Outlook attention. For each spatial location (i, j), Outlook attention calculates the similarity between it and all neighboring features in a local window of size $K*K$ centered on (i, j). Unlike self-attention, which requires query key matrix multiplication to calculate attention (i.e., $Softmax(Q^T K/\sqrt{d})$), Outlook attention simplifies this process through a reshaping operation.

Normally, when we give input$X$, we first project each $C - dim$ token, and then use two linear weights $W_A \in R^{C\times K^4}$ and $W_V \in R^{C\times C}$, respectively, and the projection is the outlook weight $A \in R^{H\times W\times K^4}$ and the value represents $V \in R^{H\times W\times C}$. $V_{\Delta_{i,j}} \in R^{C\times K^2}$ denotes all the values in the local window centered on (i, j), i.e.,

$$V_{\Delta_{i,j}} = \{V_{i+p-|\frac{K}{2}|, j+q-|\frac{K}{2}|}\}, 0 \le p, q < K \quad (3)$$

Outlook attention: The outlook weight of location (i, j) is directly used as the attention weight of value aggregation, which is reshaped into $\hat{A}_{i,j} \in MatMul(Softmax(\hat{A}_{i,j}), V_{\Delta_{i,j}})$, and then the *Softmax* function is used. Therefore, the value projection process can be written as:

$$Y_{\Delta_{i,j}} = \sum_{0\le m,n<K} Y^{i,j}_{\Delta\ i+m-|\frac{K}{2}|, j+n-|\frac{K}{2}|} \quad (4)$$

Dense aggregation: Outlook attention intensively gathers the expected value representatives, summing up weighted values from the same position of different local windows to get the output:

$$\tilde{Y}_{i,j} = \sum_{0\le m,n<K} Y^{i,j}_{\Delta\ i+m-|\frac{K}{2}|, j+n-|\frac{K}{2}|} \quad (5)$$

The implementation of the multi-head outlook attention mechanism is as follows: Assuming that the number of heads is set to N, we only need to adjust the weight shape of W_A to make$W_A \in R^{C\times N\bullet K^4}$. Then the foreground weight and value embedding are evenly divided into N segments to obtain $A_n \in R^{H\times W\times K^4}$ and $V_n \in R^{H\times W\times C_N}$, $\{n = 1, 2,..., N\}$, where the size of each head of CN satisfies $CN \times N = C$. For each (An, Vn) pair, the foreground attention is calculated separately, and then connected as the output of the multi-head foreground attention. In our study, due to the limitation of computer hardware (GPU memory only supports VOLO-D1), we mainly used VOLO-D1 to conduct the lettuce variety identification of **SP** with 384-size input images. For detailed information about several variants of the VOLO algorithm (see **Table 1**).

After constructing the lettuce variety identification model at the SP stage by using classification and detection methods, respectively, we thought about how to avoid background

TABLE 1 | Architecture information of different variants of VOLO.

Specification	VOLO-D1	VOLO-D2	VOLO-D3	VOLO-D4	VOLO-D5
Patch embedding	**8 × 8**	8 × 8	8 × 8	8 × 8	8 × 8
Stage 1 (28 × 28)	**[Head:6, stide:2 Kernel: 3 × 3 Mlp:3, dim:192] × 4**	[Head:8, stide:2 Kernel: 3 × 3 Mlp:3, dim:256] × 6	[Head:8, stide:2 Kernel: 3 × 3 Mlp:3, dim:256] × 8	[Head:12, stide:2 Kernel: 3 × 3 Mlp:3, dim:384] × 8	[Head:12, stide:2 Kernel: 3 × 3 Mlp:4, dim:384] × 12
Patch embedding	**2 × 2**	2 × 2	2 × 2	2 × 2	2 × 2
Stage 2 (14 × 14)	**[#heads:12, Mlp:3, dim:384] × 14**	[#heads:16, Mlp:3, dim:512] × 18	[#heads:16, Mlp:3, dim:512] × 28	[#heads:16, Mlp:3, dim:768] × 28	[#heads:16, Mlp:4, dim:768] × 36
Total layers	**18**	24	36	36	48
Parameters	**26.6M**	58.7M	86.3M	193M	296M

Bold highlights the specific method used in the manuscript.

interference and the similarity between plants at the same time. Finally, we propose a method, namely YOLO-VOLO. The core idea of this method is to combine the advantages of detection and classification methods to simplify the problem of plant population classification into an individual classification problem.

Step 1: we take advantage of the strong detection ability of YOLO-V5 to cut out different varieties of lettuce plants. Because there are individual growth differences between different plants, we use border filling to ensure that each picture maintains the original horizontal and vertical ratio.

Step 2: we take advantage of the strong classification ability of VOLO-D1 to construct a classification model for the individual plant images obtained in the **Step 1**.

Result Evaluation

The verification of model performance is very important. When the data amount of various samples in the training dataset is evenly distributed, the commonly used *Accuracy* is used to evaluate the performance of the model; when the data amount of various samples in the training dataset is not uniformly distributed, it is necessary to refer to other indicators to evaluate the model performance, such as *Precision*, *Recall*, and *F1-Score*. The specific definitions are as follows:

Accuracy: This is defined as the ratio of correctly classified images to the total number of lettuce images.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (6)$$

Precision: This is defined as the average of the total number of images of correctly identified lettuce varieties and the total number of images of correctly and incorrectly identified lettuce varieties.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

Recall: This is defined as the average of the images of correctly identified varieties of lettuce and the total number of correct and undetected images.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

F1-score: This is defined as the weighted average of Precision and Recall.

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (9)$$

where *TP, FP, FN*, and *TN* represent true positive, false positive, false negative and true negative, respectively.

RESULTS

Variety Recognition of Lettuce in Multiple Growth Stages Based on VOLO-D1

In order to explore the changing laws of lettuce variety identification at different growth stages, the state-of-the-art target recognition method VOLO was used to conduct a study on the variety identification of five lettuce at different growth stages. After trying different VOLO pre-training models, VOLO-D1 was finally selected as the main research method due to the limitation of computer hardware (insufficient GPU memory). The Accuracy, Recall, Precision, and F1-score are used as validation indicators to compare the classification models between different growth stages. The specific results are shown in **Table 2**.

The results show that the performance of the VOLO-D1 method in the growth stage model, except the SP stage, is close to 100%, while the model accuracy in the SP stage is only 78.381. After analysis, it is found that the dataset at the SP stage has problems such as small plant targets and large background interference, which is not conducive to accurate identification of lettuce varieties. Therefore, the classification performance of the SP stage model needs to be further improved and optimized.

Variety Recognition of Lettuce During SP Period Based on YOLO-v5

Aiming at the challenges of the SP stage dataset, the feasibility of identifying lettuce varieties through the target detection method was explored. The most advanced algorithm YOLO-v5 of the current YOLO series was used to detect and classify 5 lettuce varieties in the SP stage (as shown in **Figure 8**), the input image size remains consistent with the VOLO-D1(384*384). The specific results are shown in **Table 3**.

By comparing the result of YOLO-v5 and VOLO-D1, the results show that the F1-score of the two are relatively close (0.879 and 0.844). At the same time, YOLO-v5 is better than VOLO-D1 on the F1-score, which to a certain extent shows that the classification performance of lettuce varieties can be improved by removing background interference or increasing the number of training targets.

As we can see from **Figure 8**, the results show that in the training and verification process of the model, all the curves have converged, eliminating the possibility of model overfitting.

Variety Recognition of Lettuce During SP Period With a Fusion of YOLO and VOLO

Based on the foregoing attempts, the results prove that a single target classification and target detection method is not the best choice. Therefore, we explore the feasibility of the YOLO-VOLO-LS method proposed in this study by trying to combine the advantages of target detection and target classification algorithms.

The first step is to use the powerful detection capabilities of YOLO-V5 to uniformly modify the individual labels of the five varieties of lettuce plants to "plant," and then train the lettuce detection model. According to the lettuce coordinates output by the trained model, the lettuce plants of different varieties are cut, respectively. Due to the individual growth differences between different plants, we use border padding to ensure that each image maintains the original aspect ratio. The specific results are shown in **Table 4** and **Figure 9**.

TABLE 2 | Classification model results of different growth stages of lettuce based on VOLO-D1.

Class	Image-size	Train-acc	Val-acc	Test-acc	Recall	Precision	F1-score
SP	**384**	**99.661**	**81.970**	**78.381**	**82.946**	**85.902**	**84.398**
Day 1	384	99.999	100	100	100	100	100
Day 6	384	99.999	100	100	100	100	100
Day 12	384	99.920	100	100	100	100	100
Day 18	384	99.981	100	100	99.783	100	99.889
Day 24	384	99.999	100	100	100	100	100
Day 30	384	99.999	100	100	100	100	100

Bold highlights the results of model comparison.

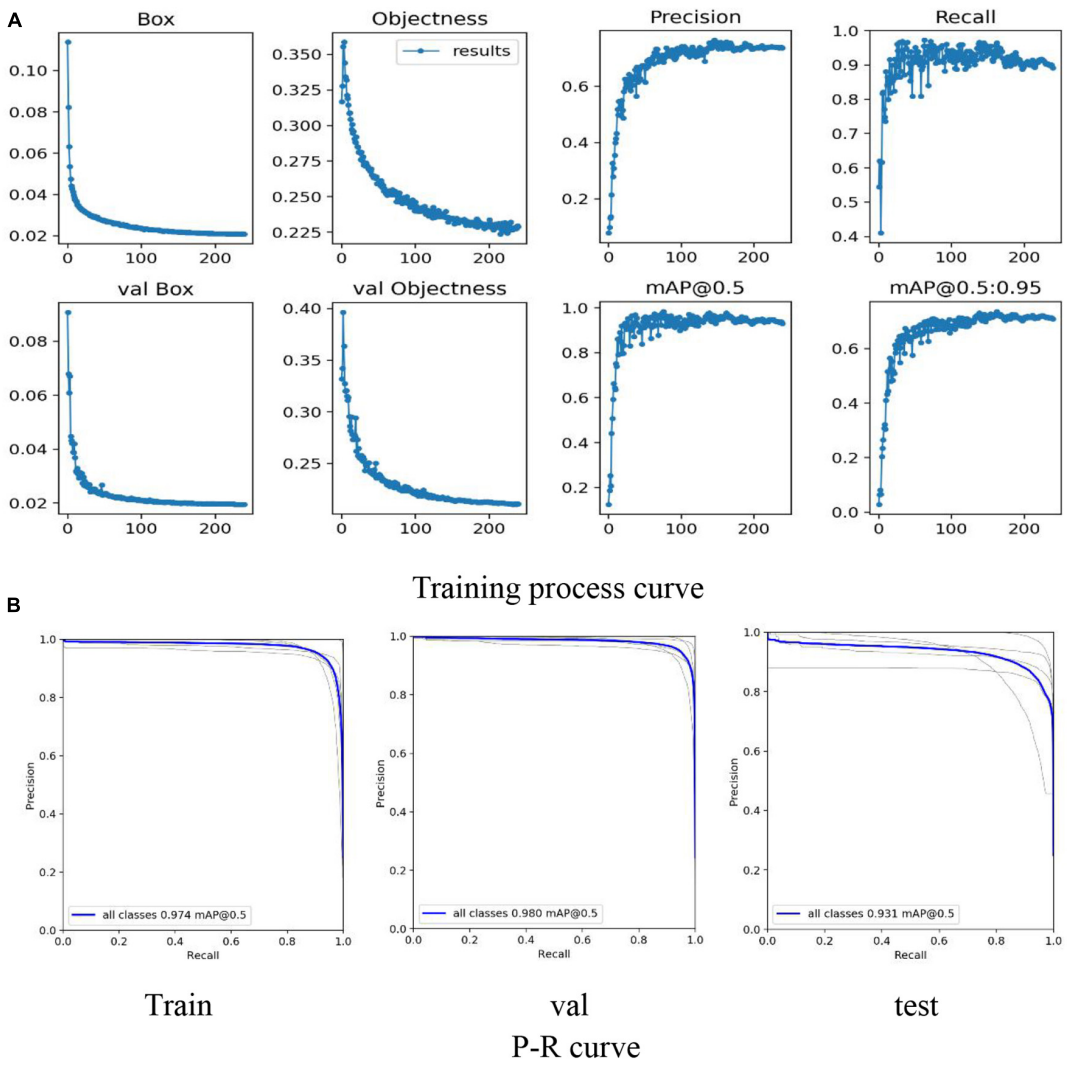


FIGURE 8 | Classification model training process based on YOLO-V5. **(A)** Mainly reflects the Loss, Precision, Recall, mAP curves in the training process and the Loss curve in the verification process. **(B)** Mainly reflects the P-R curve of training, testing and verification to judge the pros and cons of the model.

By comparing **Table 4** with **Table 3**, the result shows that when the 5-class detection task is simplified to a single-class detection task, the Recall, Precision, F1-score, mAP@0.5 are significantly improved. Among them, Recall, Precision, F1-score, and Map@0.5 have improved 0.103, 0.197, 0.157, and 0.065, respectively. Compared with the 5-class detection problem, the single-class detection task does not have the influence of problems such as the similarity between classes, so the model performance is excellent.

As shown in **Figure 9A**, all the curves in the training process have good convergence (faster convergence), and the curve is smooth, with almost no oscillations. As shown in **Figure 9B**,

TABLE 3 | Classification model results of SP stage of lettuce based on YOLO-v5.

Class	Image-size	Recall	Precision	F1-score	mAP@0.5	mAP@0.5:0.95
Train	384	0.976	0.799	0.879	0.974	0.719
Val	384	0.971	0.760	0.853	0.980	0.740
Test	384	**0.892**	**0.737**	**0.807**	0.931	0.709

Bold highlights the results of model comparison.

TABLE 4 | Detection model results of SP stage of lettuce based on YOLO-v5.

Class	Image-size	Recall	Precision	F1-score	mAP@0.5	mAP@0.5:0.95
Train	384	0.999	0.919	0.957	0.997	0.821
Val	384	0.997	0.936	0.966	0.997	0.743
Test	384	**0.995**	**0.934**	**0.964**	**0.996**	**0.694**

Bold highlights the results of model comparison.

the result shows that mAP@0.5 has a significant performance improvement, and all reach more than 0.99. Therefore, the lettuce target detection model based on YOLO-V5 can accurately detect the location of the plant, provide accurate coordinate information for the cutting process, and obtain the lettuce individual plant dataset required by the subsequent classification model.

In the second step, using the individual lettuce dataset obtained from the first steps, we further explore the classification performance of the VOLO-D1 method on this dataset. Because in the SP stage there are certain similarities between different varieties of lettuce plants, and in order to further improve the performance of the model, we used the label smoothing (LS) trick to further optimize the YOLO-VOLO model. In addition, in order to prove that reducing background interference has a greater contribution to the model than increasing the amount of data, we performed a 5*5 slicing operation on the original dataset to ensure that the amount of data in the method proposed in this study is consistent, and then use the VOLO-LS method to perform Model training. The specific results are shown in **Table 5**.

By comparing **Tables 4–6**, the F1-score was selected as the indicator to comprehensively evaluate the performance of the model. The results show that YOLO-VOLO-LS is better than VOLO-D1 and YOLO-V5 by 11.641, 15.339 on F1-score, respectively. In addition, compared with YOLO-VOLO, the results show that YOLO-VOLO-LS has increased 1.451, 3.274, 2.982, and 3.131 in terms of Test-acc, Recall, Precision, and F1-score, respectively. Compared with Slice-VOLO-LS, YOLO-VOLO-LS has increased 5.77, 8.134, 7.69, and 7.915 in terms of Test-acc, Recall, Precision, and F1-score, respectively. After analysis, by combining the advantages of target detection and target classification, not only the background interference is reduced, but also the amount of training target data is increased. Therefore, YOLO-VOLO performed well.

Comparison of Modeling Methods for Lettuce Variety Recognition

In order to further prove the performance of our proposed method YOLO-VOLO-LS, we compared the model results with mainstream classification methods such as VGG, ResNet, DenseNet, MobileNet, ShuffleNet, EfficientNet, etc. The specific results are shown in **Table 7**.

VGGNet (Simonyan and Zisserman, 2018): VGGNet is a deep CNN proposed in the early stage. Its author is a researcher from Oxford University Computer Vision Group and Google Debug. This method explores the relationship between network depth and model performance by repeatedly stacking 3*3 small convolution kernels and 2*2 maximum pooling layers, and a volume of 16–19 layer CNN is constructed. VGGNet won the runner-up of the ILSVRC 2014 competition and is the champion of the positioning project, with an error rate of 7.5% on the top 5. So far, VGGNet is still used by downstream tasks such as detection and segmentation to extract image features.

ResNet (He et al., 2016): The ResNet network is formed by adding residual units through a short-circuit mechanism on the basis of the VGG19 network. Compared with the VGG19 network, the main change of the ResNet network is to directly use the convolution of stride = 2 for downsampling, and use the global average pool layer to replace the fully connected layer. The key design principle of ResNet is that the number of feature maps is doubled when the size of the feature map is reduced by half, which maintains the complexity of the network layer. On the basis of the ResNet 18 network, ResNet 34, ResNet 50, ResNet 101, and ResNet 152 have also been proposed.

DenseNet (Huang et al., 2017): DenseNet is a CNN with dense connections between any two layers. The input of any layer of the network is the union of the outputs of all the previous layers. Unlike VGG and Inception, which improve the model in depth and width, respectively, this method starts with features and makes full use of the features of each layer in the network to achieve better model effect and fewer parameters. Therefore, the network not only strengthens the delivery and utilization of features, but also alleviates the influence of gradient disappearance during training process.

MobileNet (Howard et al., 2017): MobileNet is a lightweight CNN for embedded intelligent devices. The basic module of the network is the depthwise separable CNN, and then the lightweight network is designed based on the streamlined architecture. Among these, different convolution kernels are used for feature extraction for each input channel through depthwise revolution, and then 1 * 1 convolution check input is used for feature extraction through pointwise revolution, and then the features of the above two steps are fused. In essence, it is similar to the operation process of a standard convolution, but the amount of parameters is greatly reduced. Compared with other popular network models on ImageNet classification, MobileNet shows strong performance.

ShuffleNet (Zhang X. et al., 2018): ShuffleNet is a highly efficient CNN architecture specially applied to computer equipment with limited computing power. The architecture uses point-by-point group convolution and channel shuffling operations to use more feature mapping channels within a given computational complexity budget, so as to greatly reduce the amount of calculation while maintaining similar accuracy to the existing advanced models.

EfficientNet (Tan and Le, 2019): EfficientNet is a kind of network similar to VGG11-19, ResNet 18–101, wide-resnet 50,

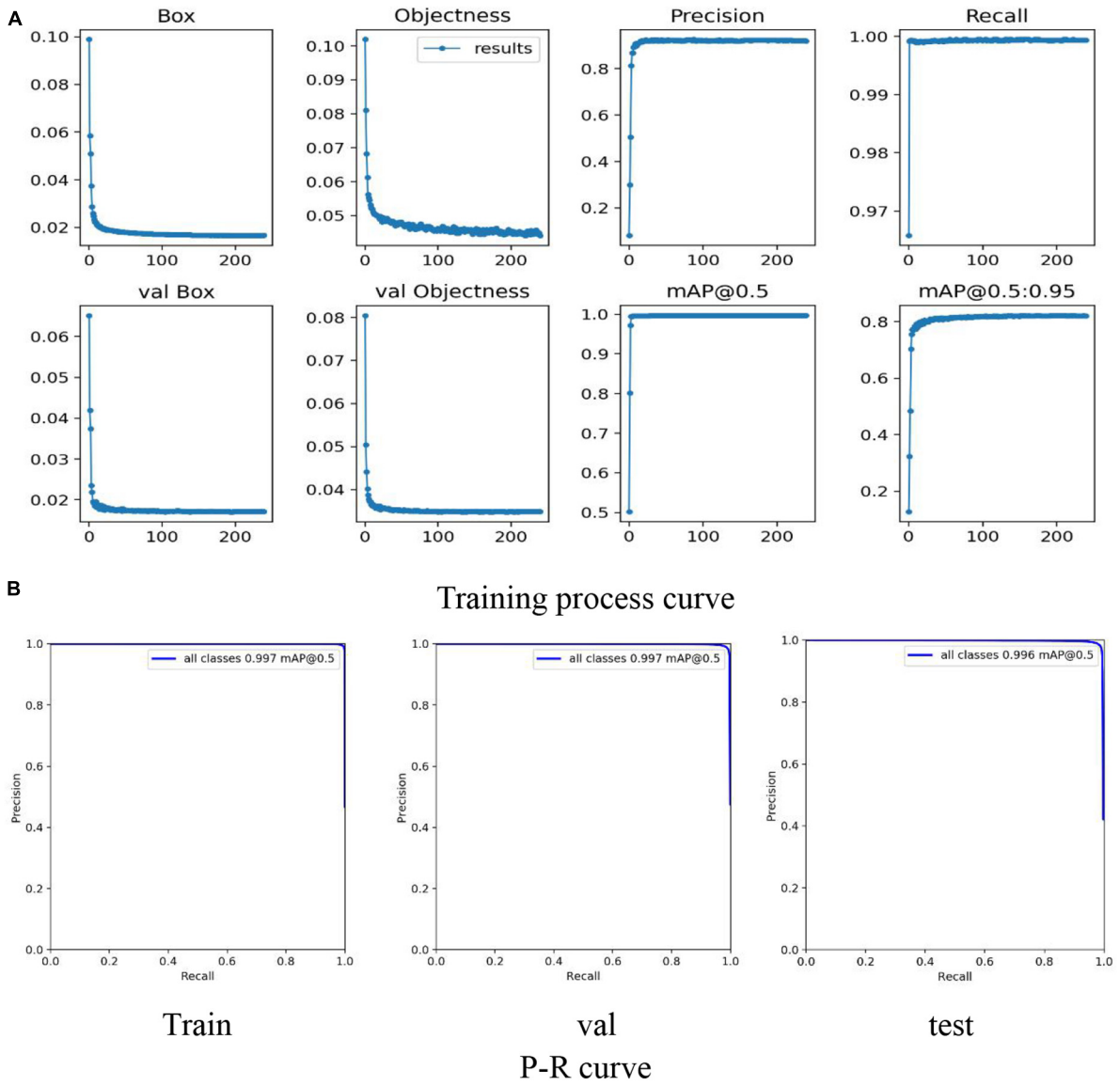


FIGURE 9 | Detection model training process based on YOLO-V5. **(A)** Mainly reflects the Loss, Precision, Recall, mAP curves in the training process and the Loss curve in the verification process. **(B)** Mainly reflects the P-R curve of training, testing and verification to judge the pros and cons of the model.

101 networks but different from those proposed by Tan and Le (2019) This network does not arbitrarily scale network dimensions such as depth, width, and resolution like traditional methods, but uses a new model scaling method that uses a series of fixed scale scaling factors to uniformly scale the network dimensions. Through the author's unremitting efforts and innovation, there are 8 types of networks: EfficentNet-b0, EfficentNet-b1, EfficentNet-b2, EfficentNet-b3, EfficentNet-b4, EfficentNet-b5, EfficentNet-b6, and EfficentNet-b7.

By comparing and analyzing the method proposed by this research with the current mainstream target classification methods, the result shows that the method proposed by this study has significant advantages in Val-acc, Test-acc, Recall, Precision, F1-score, and can effectively solve the problem of classification of lettuce varieties in the SP stage. Based on the similarity between different varieties of lettuce plants in the SP stage, the use of the LS trick also significantly improves the recognition performance of the model.

DISCUSSION

Differences in Identification of Lettuce Varieties at Different Growth Stages

In the research process of crop classification, most research mainly focuses on field crop planting scenarios, such as region type, corn, soybeans, barley, spring wheat, dried beans, sugar beets, and alfalfa, mostly supported by remote sensing technology

TABLE 5 | Classification model results of SP stage of lettuce based on YOLO-VOLO.

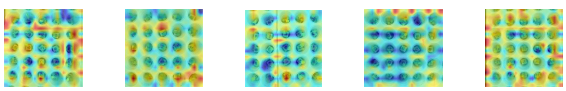
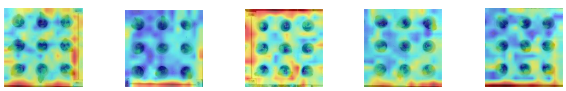
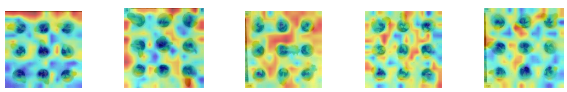
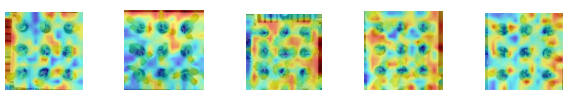
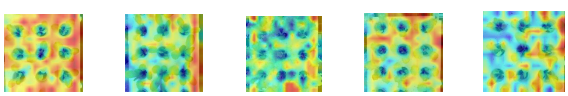
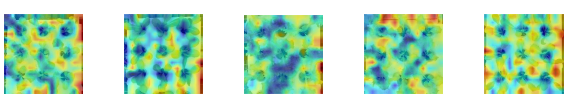
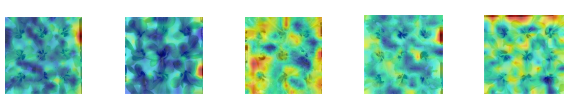
Class	Image-size	Train-acc	Val-acc	Test-acc	Recall	Precision	F1-score
YOLO-VOLO	384	99.184	**92.547**	**92.001**	**92.785**	**93.032**	**92.908**
Slice-VOLO-LS	384	99.365	87.695	87.682	87.925	88.324	88.124
YOLO-VOLO-LS	384	99.654	**95.961**	**93.452**	**96.059**	**96.014**	**96.039**

Bold highlights the results of model comparison.

(Zhang H. et al., 2018; Teimouri et al., 2019; Gadiraju et al., 2020; Mazzia et al., 2020; Sun et al., 2020). A few studies have explored crop identification methods from the perspective of leaves or veins, and most of them are supported by visual technology (Grinblat et al., 2016; Bisen, 2021). This research mainly uses the facility greenhouse as the main research scene, combined with deep learning and visual technology to explore the classification methods of small groups of crops near the ground. Therefore, we took 5 kinds of hydroponic lettuce as the research object, the VOLO-D1 method was used to construct a variety classification model for lettuce in different growth stages (see section "Variety Recognition of Lettuce in Multiple Growth Stages Based on VOLO-D1"). The results show that the recognition effect in the SP stage needs to be improved, and the recognition effect in the growth stage after transplanting is very good. In order to further analyze the reasons for this difference, we randomly obtained an image for the lettuce dataset of different growth stages to generate a Class Activation Map for analysis, and the specific results are shown in **Table 6**.

The result shows that in the SP stage, lettuce plants are small and background interference is large, and most of the attention in the model learning process is background features. After the lettuce is transplanted, during the growth stage, the plants gradually grow, and the individual differences between different varieties are gradually obvious. In addition, as the plant grows, the learning focus of the model gradually shifts from the background to the leaves of the plant, and the interference of the background on the identification of different varieties of lettuce is gradually reduced. The stronger the ability to learn key features, the better the performance of the model phenotype. Facts have proved that the crop recognition classification model with leaves as input is more effective, and this has also been indirectly proved in previous studies (Lee et al., 2017). Therefore, in view of the difficulties in the precise identification of lettuce in the SP stage, we combined the advantages of the target detection mechanism and the target classification mechanism, and we propose a new method of YOLO-VOLO-LS to solve this key problem.

TABLE 6 | The class activation map (CAM) of VOLO-D1.

Class	V1	V2	V3	V4	V5
SP					
Day 1					
Day 6					
Day 12					
Day 18					
Day 24					
Day 30					

Selection of Identification Methods for Lettuce Varieties

In the process of constructing the classification model, different methods have different advantages. By comparing the current mainstream target classification methods, the results prove that the method we propose has obvious advantages regardless of the performance of the model itself or the learning focus of the model. In order to further analyze the difference in model performance between different methods, we use the Class Activation Map method to analyze the learning focus of different models, and the specific results are shown in **Table 8**. The result shows that different methods focus on different points in the model training process. Some methods can only extract part of the image features during the training process. For example, VGG and ResNet mainly extract the edge features of the image. Among them, VGG16 replaces the larger convolution kernel with a continuous 3 * 3 convolution kernel while increasing the network depth (e.g., 11 * 11, 7 * 7, 5 * 5), under the given receptive field conditions, the stacking effect of small convolution kernel is better than that of large convolution kernel (Simonyan and Zisserman, 2018). ResNet adds a direct channel between layers of the network, which effectively avoids the loss of information transmission between layers and reduces the possibility of gradient disappearance or gradient explosion (He et al., 2016). Therefore, ResNet is better than VGG in feature information extraction and retention. Some methods can only extract the central and local features of the image during the training process, such as DenseNet169, MobileNet_v2, ShuffleNet_v2, EfficientNet-B4 mainly focus on the central area of the image, and the edge feature information is lost. Among

TABLE 7 | Comparison of modeling methods for lettuce variety recognition.

Class	Image-size	Train-acc	Val-acc	Test-acc	Recall	Precision	F1-score
VGG16	384	99.863	85.861	77.562	86.580	87.685	87.129
ResNet50	384	99.782	73.872	73.846	72.124	74.137	73.117
DenseNet169	384	99.736	81.975	77.254	81.973	85.387	83.645
MobileNet_v2	384	99.936	74.395	70.872	72.108	80.921	76.261
ShuffleNet_v2	384	97.826	72.414	72.441	73.631	78.675	76.069
EfficientNet-B4	384	99.563	73.128	72.340	74.123	85.348	79.341
YOLO-VOLO	384	99.181	**92.542**	**92.001**	**92.785**	**93.032**	**92.908**
YOLO-VOLO-LS	384	99.652	**95.961**	**93.453**	**96.059**	**96.014**	**96.039**

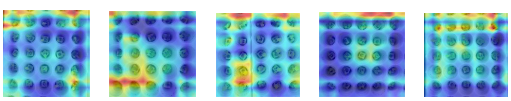
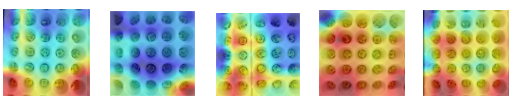
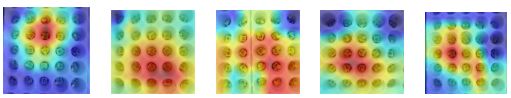
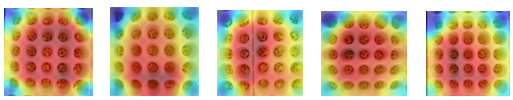
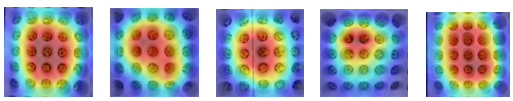
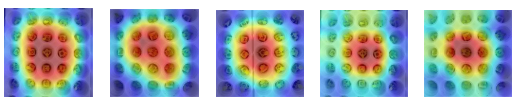
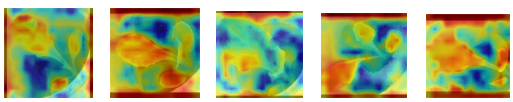
Bold highlights the results of model comparison.

them, DenseNet169 is different from the previous improvements in network length and width, and in order to make maximum use of the characteristic information between layers, DenseNet169 connects all layers on the premise of ensuring the maximum information transmission between layers, making the network narrower, making fewer parameters and producing a better effect (Huang et al., 2017). ShuffleNet_v2 uses channel splitting to achieve the effect of feature reuse, so as to improve the computational efficiency of the model (Zhang X. et al., 2018). EfficientNet-B4 improves the performance of the model mainly through a synergy coefficient in terms of network depth, width and resolution (. MobileNet_v2 based on the residual block, first uses 1 × 1 lower channel to pass through ReLu, then uses 3 × 3 space convolution to pass through ReLu, and then uses 1 × 1 convolution to recover the channel, which reduces the amount of calculation and improves the performance of the model (Howard et al., 2017). Therefore, MobileNet_v2 is superior to other methods in feature extraction. However, due to the different emphasis of each method, the performance effect of the model may be different. In the actual use process, a method suitable for your own data set is selected through comparative analysis.

In this study, a method called YOLO-VOLO-LS is proposed based on analyzing the characteristics of the lettuce dataset in the SP stage, and aimed at solving the problems of small target detection, large background interference, and high individual similarity by combining the advantages of target detection and target classification. Considering the cost of data labeling and the performance of the model, based on the respective advantages of target detection and target classification, we adopted the strategy of first detection and then classification to classify the lettuce in the SP stage. Through this process, we simplified the group target classification problem into an individual target classification problem. While minimizing the influence of the background on the classification of lettuce varieties, the model can learn more leaf details to improve the recognition ability of the model. By observing the Class Activation Maps of YOLO-VOLO-LS and other methods, we can clearly find that our proposed method can almost completely extract the characteristic information of lettuce plants, which is why this method has obvious advantages in accuracy. Similar studies have also proved that crop classification based on the characteristics of leaves or veins has a significant improvement in the model effect (Lee et al., 2017; Tan et al., 2020; Bisen, 2021). This is why our study uses first detection and then classification when classifying lettuce in the SP stage.

In addition, the method we propose plays a role of data enhancement to a certain extent. In order to verify the contribution of increasing the amount of data and reducing the background interference to the SP stage lettuce variety recognition model, we compare the results in section "Variety Recognition of Lettuce During SP Period With a Fusion of YOLO and VOLO." By slicing the original data set with 5 rows and 5 columns, the data volume can be consistent with the YOLO-VOLO-LS method, and then the VOLO-LS method is used for training verification, and it is found that the improvement of the model performance is very limited.

TABLE 8 | The class activation map (CAM) of different methods.

Class	V1	V2	V3	V4	V5
VGG16					
ResNet50					
DenseNet169					
MobileNet_v2					
ShuffleNet_v2					
EfficientNet-B4					
YOLO-VOLO-LS					

Therefore, we found that removing the background interference to the maximum extent contributes the most to the model, which further verifies the effectiveness of the method proposed in this study. At the same time, during the image slicing process, there is no guarantee that the target in the image is completely segmented, which may cause some plants to be damaged during the slicing process. This may also be the reason for the general performance of the sliced dataset.

Although the method proposed in this study is effective, it still has certain limitations. First, the method proposed in this research is more suitable for the variety identification of small target crops in low-altitude and high-density scenarios such as facility greenhouses. The early variety identification of field crops obtained from high-altitude scenes such as remote sensing has yet to be tried and verified. Second, although the method of first detection and then classification can significantly improve the early variety recognition effect of lettuce seedlings, the specific calculation process may take a long time. Finally, in the follow-up research process, on the one hand, we plan to build a set of software and hardware intelligent detection systems suitable for different growth periods for greenhouse crops based on existing research, and on the other hand, in order to further improve the applicability of this method, we want to apply this method to early crops in field scenarios.

CONCLUSION

In this study, a variety identification model was constructed for hydroponic lettuce grown in a greenhouse under the conditions of different growth periods. The results found that the performance of the lettuce variety classification model at the SP stage before the lettuce transplantation still needs to be improved. By combining the respective advantages of the target detection mechanism and the target classification mechanism, we innovatively propose a classification method for lettuce varieties at the SP stage, called YOLO-VOLO-LS. This method has achieved excellent results of 95,961, 93,452, 96,059, 96,014, 96,039 in Val-acc, Test-acc, Recall, Precision, and F1-score, respectively. In addition, we have achieved nearly 100% of the lettuce classification effect in the growth stages of Days 1, 6, 12, 18, 24, and 30 by adopting the VOLO-D1 method. In view of the characteristics of lettuce seedlings in the SP stage, we simplified the group classification problem to an individual classification problem by adopting the strategy of first detection and then classification, which significantly improved the performance of the model. Of course, this method may be more suitable for research on the variety identification of high-density small target crops in a low-altitude environment. The small target image of the group can be cropped through the detection method, which not only increases the amount of data, but also reduces the background interference. Therefore, through the combination of detection and classification methods, on the one hand, the problems of small target crop similarity and background interference can be overcome, and on the other hand, the problem of small samples can be solved to a certain extent, which has a certain contribution in data preprocessing.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

PZ collected literature and training models together and completed the manuscript writing. DL modified the manuscript. Both authors read and approved the final manuscript.

FUNDING

This work was supported by the Guolian Aquatic Products Development Co., Ltd., Program "Big data analysis and management cloud service platform construction and large-scale application of shrimp" (Grant no. 2017B010126001) and the Major Science and Technology Innovation Project of Shandong Province "Research and development of key technologies for accurate monitoring, early warning and regulation of land and sea relay water quality" (Grant no. 2019JZZY010703).

ACKNOWLEDGMENTS

We would like to thank the editor and reviewers for their valuable input, time, and suggestions to improve the quality of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.806878/full#supplementary-material

REFERENCES

Alzamily, J. Y., Salim, S., and Naser, A. (2019). Lemon classification using deep learning. *Int. J. Acad. Pedagog. Res.* 3, 16–20.

Bhosle, K., and Musande, V. (2019). Evaluation of deep learning CNN model for land use land cover classification and crop identification using hyperspectral remote sensing images. *J. Indian Soc. Remote Sens.* 47, 1949–1958.

Bisen, D. (2021). Deep convolutional neural network based plant species recognition through features of leaf. *Multimed. Tools Appl.* 80, 6443–6456. doi: 10.1007/s11042-020-10038-w

Chew, R., Rineer, J., Beach, R., O'neil, M., Ujeneza, N., Lapidus, D., et al. (2020). Deep neural networks and transfer learning for food crop identification in UAV images. *Drones* 4, 1–14. doi: 10.3390/drones4010007

Dharwadkar, S., Bhat, G., Subba Reddy, N. V., and Aithal, P. K. (2017). "Floriculture classification using simple neural network and deep learning," in *Proceedings of the RTEICT 2017 – 2nd International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) 2018-Janua*, (Piscataway, NJ: IEEE), 619–622. doi: 10.1109/RTEICT.2017.8256671

Dileep, M. R., and Pournami, P. N. (2019). "AyurLeaf: a deep learning approach for classification of medicinal plants," in *Proceedings of the IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON 2019-Octob*, (Piscataway, NJ: IEEE), 321–325. doi: 10.1109/TENCON.2019.8929394

Duong-Trung, N., Quach, L. D., Nguyen, M. H., and Nguyen, C. N. (2019). A combination of transfer learning and deep learning for medicinal plant classification. *PervasiveHealth Pervasive Comput. Technol. Healthc. Part* F1479, 83–90. doi: 10.1145/3321454.3321464

Gadiraju, K. K., Ramachandra, B., Chen, Z., and Vatsavai, R. R. (2020). "Multimodal deep learning based crop classification using multispectral and multitemporal satellite imagery," in *Proceeding of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, (New York, NY: Association for Computing Machinery), 3234–3242. doi: 10.1145/3394486.3403375

Ghazi, M. M., Yanikoglu, B., and Aptoula, E. (2016). "Open-set plant identification using an ensemble of deep convolutional neural networks," in *CEUR Workshop Proceedings*, Évora, Vol. 1609, 518–524.

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

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016-Decem* (Piscataway, NJ: IEEE), 770–778. doi: 10.1109/CVPR.2016.90

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," in *Proceedings of the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*. Available online at: http://arxiv.org/abs/1704.04861 (accessed June, 2020)

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Piscataway, NJ: IEEE), 2261–2269. doi: 10.1109/CVPR.2017.243

Jia, W., Xu, S., Liang, Z., Zhao, Y., Min, H., Li, S., et al. (2021). Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector. *IET Image Process.* 15, 3623–3637. doi: 10.1049/ipr2.12295

Kasper-eulaers, M., Hahn, N., Berger, S., Sebulonsen, T., and Myrland, Ø. (2021). Short communication: detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5. *Algorithms* 14:114.

Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., and Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. *Comput. Electron. Agric.* 158, 20–29. doi: 10.1016/j.compag.2019.01.041

Khamparia, A., Singh, A., Luhach, A. K., Pandey, B., and Pandey, D. K. (2020). Classification and identification of primitive kharif crops using supervised deep convolutional networks. *Sustain. Comput. Inform. Syst.* 28:100340. doi: 10.1016/j.suscom.2019.07.003

Kim, Y., Kwak, G.-H., Lee, K.-D., Na, S.-I., Park, C.-W., and Park, N.-W. (2018). 작물분류에서 기계학습 및 딥러닝 알고리즘의 분류 성능 평가: 하이퍼파라미터 튜닝과 훈련자료 크기의 영향 분석 Performance evaluation of machine learning and deep learning algorithms in crop classification?: impact of hyper-parameters and training sample size. *Korean J. Remote Sens.* 34, 811–827. doi: 10.7780/kjrs.2018.34.5.9

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* 14, 778–782. doi: 10.1109/LGRS.2017.2681128

Le, T. T., Lin, C. Y., and Piedad, E. J. (2019). Deep learning for noninvasive classification of clustered horticultural crops – A case for banana fruit tiers. *Postharvest Biol. Technol.* 156:110922. doi: 10.1016/j.postharvbio.2019.05.023

Lee, S. H., Chan, C. S., Mayo, S. J., and Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit.* 71, 1–13. doi: 10.1016/j.patcog.2017.05.015

Liu, C., Lu, W., Gao, B., Kimura, H., Li, Y., and Wang, J. (2020). Rapid identification of chrysanthemum teas by computer vision and deep learning. *Food Sci. Nutr.* 8, 1968–1977. doi: 10.1002/fsn3.1484

Liu, F., Ren, X., Zhang, Z., Sun, X., and Zou, Y. (2021). "Rethinking skip connection with layer normalization," in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, 3586–3598. doi: 10.18653/v1/2020.coling-main.320

Liu, W., Wang, Z., Zhou, B., Yang, S., and Gong, Z. (2021). Real-time signal light detection based on yolov5 for railway. *IOP Conf. Ser. Earth Environ. Sci.* 769:042069. doi: 10.1088/1755-1315/769/4/042069

Liu, Z., Wang, J., Tian, Y., and Dai, S. (2019). Deep learning for image-based large-flowered chrysanthemum cultivar recognition. *Plant Methods* 15, 1–11. doi: 10.1186/s13007-019-0532-7

Mazzia, V., Khaliq, A., and Chiaberge, M. (2020). Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-Convolutional Neural Network (R-CNN). *Appl. Sci.* 10:238. doi: 10.3390/app10010238

Mei, X., Nie, W., Liu, J., and Huang, K. (2018). "Polsar image crop classification based on deep residual learning network," in *Proceedings of the 2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics)* (Piscataway, NJ: IEEE), 1–6. doi: 10.1109/Agro-Geoinformatics.2018.8476061

Nkemelu, D. K., Omeiza, D., and Lubalo, N. (2018). Deep convolutional neural network for plant seedlings classification. *arXiv* [preprint]. arxiv:1811.08404,

Pereira, C. S., Morais, R., and Reis, M. J. C. S. (2019). Deep learning techniques for grape plant species identification in natural images. *Sensors (Switzerland)* 19:4850. doi: 10.3390/s19224850

Rong, D., Wang, H., Ying, Y., Zhang, Z., and Zhang, Y. (2020). Peach variety detection using VIS-NIR spectroscopy and deep learning. *Comput. Electron. Agric.* 175:105553. doi: 10.1016/j.compag.2020.105553

Simonyan, K., and Zisserman, A. (2018). VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION Karen. *Am. J. Heal. Pharm.* 75, 398–406.

Suh, H. K., IJsselmuiden, J., Hofstee, J. W., and van Henten, E. J. (2018). Transfer learning for the classification of sugar beet and volunteer potato under field conditions. *Biosyst. Eng.* 174, 50–65. doi: 10.1016/j.biosystemseng.2018.06.017

Sun, Y., Liu, Y., Wang, G., and Zhang, H. (2017). Deep learning for plant identification in natural environment. *Comput. Intell. Neurosci.* 2017:7361042. doi: 10.1155/2017/7361042

Sun, Z., Di, L., Fang, H., and Burgess, A. (2020). Deep learning classification for crop types in north dakota. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 13, 2200–2213. doi: 10.1109/JSTARS.2020.2990104

Tan, J. W., Chang, S. W., Abdul-Kareem, S., Yap, H. J., and Yong, K. T. (2020). Deep learning for plant species classification using leaf vein morphometric. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 17, 82–90. doi: 10.1109/TCBB.2018.2848653

Tan, M., and Le, Q. V. (2019). "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning. ICML 2019 2019-June*, Long Beach, CA, 10691–10700.

Teimouri, N., Dyrmann, M., and Jørgensen, R. N. (2019). A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. *Remote Sens.* 11, 1–18. doi: 10.3390/rs11080893

Tiwari, S. (2020). A comparative study of deep learning models with handcraft features and non-handcraft features for automatic plant species identification. *Int. J. Agric. Environ. Inf. Syst.* 11, 44–57. doi: 10.4018/IJAEIS.2020040104

Tóth, B. P., Osváth, M., Papp, D., and Szucs, G. (2016). Deep learning and SVM classification for plant recognition in content-based large scale image retrieval. *CEUR Workshop Proc.* 1609, 569–578.

Van Hieu, N., and Hien, N. L. H. (2020). Automatic plant image identification of Vietnamese species using deep learning models. *Int. J. Eng. Trends Technol.* 68, 25–31. doi: 10.14445/22315381/IJETT-V68I4P205S

Yalcin, H., and Razavi, S. (2016). "Plant classification using convolutional neural networks," in *Proceedings of the 2016 5th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2016*, (Piscataway, NJ: IEEE). doi: 10.1109/Agro-Geoinformatics.2016.7577698

Yoosefzadeh-Najafabadi, M., Tulpan, D., and Eskandari, M. (2021). Application of machine learning and genetic optimization algorithms for modeling and optimizing soybean yield using its component traits. *PLoS One* 16:e0250665. doi: 10.1371/journal.pone.0250665

Yuan, L., Hou, Q., Jiang, Z., Feng, J., and Yan, S. (2021). "VOLO: vision outlooker for visual recognition," in *Proceedings of the IEEE Computer Society Conference*

on Computer Vision and Pattern Recognition. Available online at: http://arxiv.org/abs/2106.13112 (accessed June, 2021).

Zhang, B., Zhao, L., and Zhang, X. (2020). Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sens. Environ.* 247:111938. doi: 10.1016/j.rse.2020.111938

Zhang, H., He, G., Peng, J., Kuang, Z., and Fan, J. (2018). "Deep learning of path-based tree classifiers for large-scale plant species identification," in *Proceedings of the – IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018* (Piscataway, NJ: IEEE), 25–30.

Zhang, X., Zhou, X., Lin, M., and Sun, J. (2018). "ShuffleNet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 6848–6856. doi: 10.1109/CVPR.2018.00716

Zhang, Z., Liu, H., Meng, Z., and Chen, J. (2019). Deep learning-based automatic recognition network of agricultural machinery images. *Comput. Electron. Agric.* 166:104978. doi: 10.1016/j.compag.2019.104978

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 March 2022
doi: 10.3389/fpls.2022.834447

Enhancing Wheat Disease Diagnosis in a Greenhouse Using Image Deep Features and Parallel Feature Fusion

Zhao Zhang[1,2], Paulo Flores[3]*, Andrew Friskop[4], Zhaohui Liu[4], C. Igathinathane[3], X. Han[5,6], H. J. Kim[6,7], N. Jahan[3], J. Mathew[3] and S. Shreya[8]*

[1] Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing, China, [2] Key Lab of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing, China, [3] Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo, ND, United States, [4] Department of Plant Sciences, North Dakota State University, Fargo, ND, United States, [5] Department of Biosystems Engineering, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon, South Korea, [6] Interdisciplinary Program in Smart Agriculture, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon, South Korea, [7] Department of Biosystems and Biomaterials Engineering, College of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea, [8] Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND, United States

OPEN ACCESS

Edited by:
Yiannis Ampatzidis, University of Florida, United States

Reviewed by:
Zhou Zhang, University of Wisconsin-Madison, United States
Yufei Liu, Zhejiang University, China
Prabira Kumar Sethy, Sambalpur University, India
Yaqoob Majeed, University of Agriculture, Faisalabad, Pakistan

***Correspondence:**
Zhao Zhang
zhaozhangcau@cau.edu.cn
Paulo Flores
paulo.flores@ndsu.edu

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *13 December 2021*
Accepted: *27 January 2022*
Published: *10 March 2022*

Citation:
Zhang Z, Flores P, Friskop A, Liu Z, Igathinathane C, Han X, Kim HJ, Jahan N, Mathew J and Shreya S (2022) Enhancing Wheat Disease Diagnosis in a Greenhouse Using Image Deep Features and Parallel Feature Fusion. Front. Plant Sci. 13:834447. doi: 10.3389/fpls.2022.834447

Since the assessment of wheat diseases (e.g., leaf rust and tan spot) *via* visual observation is subjective and inefficient, this study focused on developing an automatic, objective, and efficient diagnosis approach. For each plant, color, and color-infrared (CIR) images were collected in a paired mode. An automatic approach based on the image processing technique was developed to crop the paired images to have the same region, after which a developed semiautomatic webtool was used to expedite the dataset creation. The webtool generated the dataset from either image and automatically built the corresponding dataset from the other image. Each image was manually categorized into one of the three groups: control (disease-free), disease light, and disease severity. After the image segmentation, handcrafted features (HFs) were extracted from each format of images, and disease diagnosis results demonstrated that the parallel feature fusion had higher accuracy over features from either type of image. Performance of deep features (DFs) extracted through different deep learning (DL) models (e.g., AlexNet, VGG16, ResNet101, GoogLeNet, and Xception) on wheat disease detection was compared, and those extracted by ResNet101 resulted in the highest accuracy, perhaps because deep layers extracted finer features. In addition, parallel deep feature fusion generated a higher accuracy over DFs from a single-source image. DFs outperformed HFs in wheat disease detection, and the DFs coupled with parallel feature fusion resulted in diagnosis accuracies of 75, 84, and 71% for leaf rust, tan spot, and leaf rust + tan spot, respectively. The methodology developed directly for greenhouse applications, to be used by plant pathologists, breeders, and other users, can be extended to field applications with future tests on field data and model fine-tuning.

Keywords: wheat disease, plant pathology, deep features, handcrafted features, data fusion

HIGHLIGHTS

- Parallel feature fusion of different types of images improved the accuracy of wheat disease diagnosis.
- Deep features outperformed handcrafted features in wheat disease detection.
- Deep features extracted by deep-layered models produced higher accuracy.
- A free semiautomatic webtool for expedited paired dataset creation was developed and made available.

INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the world's most productive and important crops, which plays a crucial role in food security (Curtis and Halford, 2014; Shewry and Hey, 2015). Currently, wheat production faces a number of challenges, among which diseases are ranked among the top (Bolton et al., 2008). In addition to reducing yield, wheat diseases could lower the grain quality or even result in grain contamination due to toxins produced by pathogens (Lu et al., 2017; Qiu et al., 2019). Leaf rust and tan spot are common diseases that affect wheat production in the United States and worldwide, which can cause wheat yield losses of 10–40% (De Wolf, 2008; Sharma et al., 2016).

The two main approaches to manage wheat diseases are breeding disease-resistant varieties and through chemical applications (Kolmer, 1996; Ransom and McMullen, 2008; Figlan et al., 2020). For both approaches, researchers conduct extensive greenhouse work before transferring the most promising materials or treatments to the field for further evaluation. Hence, it is critical for researchers to obtain accurate information on the disease conditions in the greenhouse (Abdulridha et al., 2020). The current approach of wheat disease diagnosis relies on visual observations by well-trained graders. This approach potentially suffers from subjectivity (grader bias), inefficiency (slow speed of observation), inter-grader variation (inconsistent results among different graders), and fatigue (tiresome operation) (Lehmann et al., 2015). Therefore, an automated, efficient, and objective approach to accurately and quickly diagnose wheat diseases is needed (Luvisi et al., 2016).

Leaf rust is characterized by the presence of rust-colored pustules erupting at the crop leaves (Salgado et al., 2016). Tan spot symptoms are oval or diamond-shaped to elongated irregular spots on the leaf, and these spots enlarge and turn tan with a yellow border and a small dark brown spot near the center (McMullen and Adhikari, 2009). Many studies have taken advantage of these visible symptoms and used color [red, green, blue (RGB)] images, coupled with different classifiers, for disease detection (Johannes et al., 2017; Lu et al., 2017; Saleem et al., 2019; Su et al., 2020). Color images are the dominant type of images used for crop disease detection because of their low cost and easiness to acquire and handle (Gaikwad and Musande, 2017; Barbedo, 2018; Qiu et al., 2019; Wiesner-Hanks et al., 2019). In addition to color images, color-infrared (CIR) images have been extensively used in crop disease detection (Lehmann et al., 2015). Different from the color images consisting of RGB, CIR images include three bands, namely, near-infrared (NIR), red, and green bands. The CIR images take advantage of the fact that disease lesions (chlorotic or necrotic) cause biochemical changes on the plant tissue, which can significantly affect the energy reflection on NIR of the electromagnetic spectrum (Roberts et al., 1993; Franke and Menz, 2007; Azadbakht et al., 2019). Healthy plants on the CIR images usually display high reflectance on the NIR band and low reflectance on the red band, while an opposite band reflectance pattern is observed on the unhealthy plants (Carlson and Ripley, 1997). Based on this principle, vegetation indices can be calculated from CIR images. Among them, the Normalized Difference Vegetation Index (NDVI) has been extensively used for monitoring the crop health condition (Bravo et al., 2003; Franke and Menz, 2007; Su et al., 2018; Yang, 2020). CIR images were utilized for wheat and cotton disease detection, with practical application of the results (Bajwa and Tian, 2001, 2002; Moshou et al., 2004; Yang, 2020). Although both RGB and CIR images have been extensively used for crop disease detection, few studies were conducted to compare their performance on wheat disease detection and further improve the methodology and detection accuracy.

After collecting color or CIR images, handcrafted features (HFs), which are extracted from images using algorithms to represent the physical characteristics of the plants, would serve as the basis for classification purposes (Zhang et al., 2016, 2020b; Jahan et al., 2020). Numerous studies associated with crop disease diagnosis have been carried out based on the HFs, including vegetation indices (Ashourloo et al., 2014; Chen et al., 2018), texture (Wood et al., 2012; Sun et al., 2019; Wan et al., 2020), and color (Patil and Kumar, 2011; Gaikwad and Musande, 2017). The HFs-based classification requires domain knowledge on feature selection, as the classification accuracy mainly determines whether the selected features have a good representation of the diseases (Zhang et al., 2020a). One approach to get rid of the domain knowledge required by the HFs-based classification is through deep learning (DL).

During the last decade, DL has experienced significant progress regarding image classification, with the convolutional neural networks (CNNs) having been the core (Zhang et al., 2020b). The CNNs enabled the implementation of algorithms for automatic feature extraction, which does not require domain knowledge. Very recently, deep features (DFs; features automatically extracted by CNNs) have been used in crop disease detection, and the literature in this field is limited. Lu et al. (2017) extracted DFs and applied them to discriminate wheat diseases, such as leaf blotch, smut, stripe rust, and black chaff, but the algorithms' performance on disease severity diagnosis was not reported. In addition, DFs have been used for the detection of apple scab disease (Khan et al., 2018) and rice leaf disease (Sethy et al., 2020). However, few studies have reported the application of DFs to differentiate and assess the severity of wheat leaf rust and tan spot diseases. Furthermore, the diagnosis accuracies based on selected features (HFs and DFs), which influence classification performance, were unavailable.

Features (usually from a single-source image) were extracted and then fed into classifiers for classification (Yang et al., 2003; Sethy et al., 2020). Features extracted from color images were

used to detect tomato leaf diseases (Patil and Kumar, 2011) and corn diseases (Wiesner-Hanks et al., 2019), while CIR images were used for assessing cotton rot disease (Yang, 2020). Since a certain type of image might only provide partial information to aid plant diseases classification, researchers (Bulanon et al., 2009; Castanedo, 2013) had been experimenting with data fusion techniques by combining features from different types of images to improve the model accuracy. Color image blended with NIR image detected the freshness level of food products and demonstrated an improved classification accuracy over either single-source image (Huang et al., 2016). Integration of color and thermal images improved the field orange detection accuracy (Bulanon et al., 2009). In the color images, oranges were not well differentiated from leaves because of similarities between them. However, they had different temperatures, which were obtained *via* thermal images. Thus, the fusion of the features from color and thermal images led to a higher accuracy. Even though the use of data fusion techniques has resulted in higher classification accuracy, few studies have fused color and CIR images information for wheat disease diagnosis.

With an overall goal of developing and implementing an automated solution to assess greenhouse wheat diseases (e.g., leaf rust, tan spot, and leaf rust + tan spot), this study proposes an innovative methodology of using deep features and their parallel fusion from color and CIR images. Specific objectives of this study were: (1) to compare the performance of features from color and CIR images, and their parallel fusion in wheat disease diagnosis; (2) to compare the accuracies of DFs extracted from different DL models on wheat disease diagnosis and select the one generating the highest accuracy; and (3) to compare the accuracies of HFs and DFs in wheat disease detection.

MATERIALS AND METHODS

The various process steps followed in this study to improve wheat disease diagnosis accuracy using feature fusion and DFs are illustrated in **Figure 1**. After collecting color and CIR images for the same plants in a paired mode, the region of interest (ROI) was automatically determined using image processing techniques. We developed a webtool to expedite the dataset generation—while manually cropping either type of images (color or CIR), the corresponding image of the other type would be generated automatically (paired dataset). After generating the dataset, features (HFs and DFs) extracted from single-source images (color or CIR) and their fusion were fed into a support vector machine (SVM) for accuracy comparison. Finally, the methodology yielding the highest accuracy would be recommended for future application. The following sections describe the processes in detail.

Image Acquisition

The experiment was conducted at the North Dakota State University, Agricultural Experiment Station Research Greenhouse Complex (Fargo, ND, United States). Since the greenhouse is enclosed with a transparent roof and windows, the crop growing light conditions can be considered as semi-natural illumination. Two wheat varieties (*Prosper* for leaf rust disease and *Jerry* for tan spot disease) were planted in pots (Deepot D40: 6.4 × 25.4 cm; Stuewe and Sons, Inc., Tangent, Oregon). For the two disease groups, crops were properly inoculated and then kept in the incubation chamber for 24 h for expedited development of the diseases. The control group was kept in another incubation chamber with the same conditions and time as the disease group. After inoculation, the two diseases required different amounts of time to display symptoms—about 10 days for leaf rust and 6 days for a tan spot. Immediately after the initial observation of disease symptoms, the data collection started and continued for the next 12 consecutive days, with images being collected between 10:00 a.m. and 12:00 p.m.

Two off-the-shelf cameras were used for data collection—a Canon EOS Rebel T7i camera (Ota City, Tokyo, Japan) for color images (6,000 × 4,000 pixel resolution) and a multi-band camera (LDP LLC, Carlstadt, NJ, United States) for CIR images (5,184 × 3,456 pixel resolution) (**Figure 2**). A frame (60 × 60 × 90 cm) built to facilitate image collection was used as a reference for image collection for both cameras—the diameter of the hole at the top sheet of the frame was a little wider than the diameter of the camera lens, allowing them to go through to capture the images. A rack that could hold 8 pots was placed at the center of the frame bottom for the collection of both color and CIR images, after which the rack was replaced. That process was repeated until all plants were imaged. There were 10 racks for each variety of plants, for a total of 160 pots of plants (8 pots × 10 racks × 2 varieties).

Color and Color-Infrared Image Datasets Creation

Automatic Raw Image Cropping

After collecting color and CIR images, the next step was to prepare the paired datasets (the same portion of plants shown in both color and CIR images). A critical requirement during the paired dataset preparation was to have the color image to be corresponding to the CIR image. Since the two cameras had different field of views and resolutions, as well as the image collection positions were not exactly the same for the two cameras, the views of the two images were different, as shown in **Figures 2C,E**. It is thus necessary to keep the views of the two images the same for further paired dataset generation by proper cropping. Previous research used a manual approach for image cropping (Bulanon et al., 2009), which is inefficient and inaccurate. In this study, an automatic raw image cropping approach was developed and applied, which used the aluminum square base of the experimental frame as a reference. The frame was first detected using color thresholding, and then, the mask was generated (after noise removal) for each type of image. Only the image section within the square base was kept, and detailed procedures and parameters for image processing are shown in **Figure 3**.

Webtool for Paired Image Dataset Creation

After the paired images were automatically cropped, they covered the same view. While creating the paired dataset, it required certain plant regions to be present in the paired color and CIR

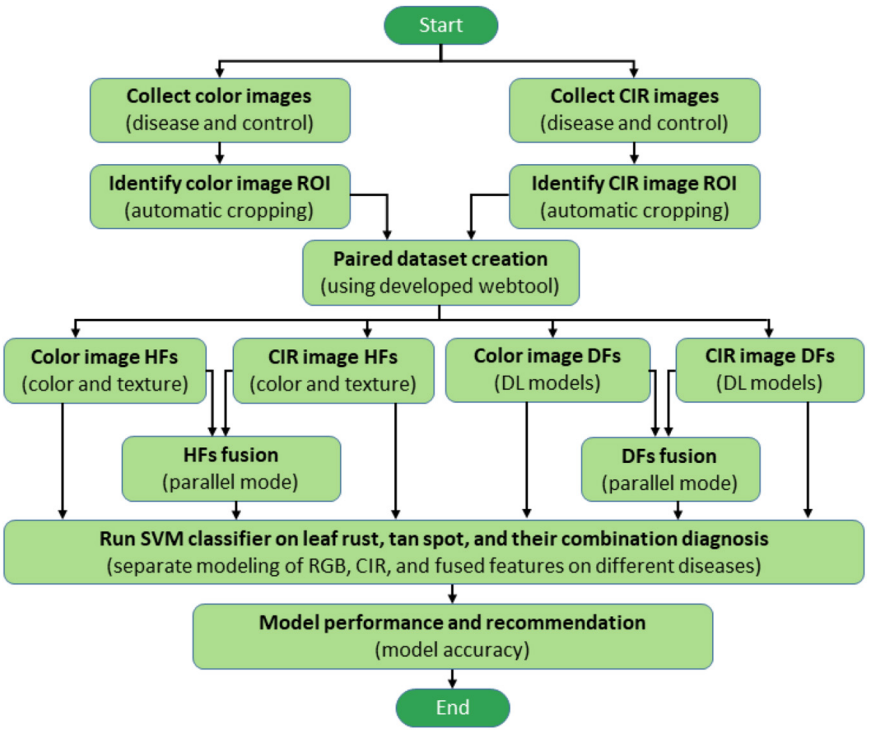


FIGURE 1 | Overall process flowchart of wheat disease diagnosis. ROI, CIR, DL, HFs, DFs, and SVM represent the region of interest, color-infrared, deep learning, handcrafted features, deep features, and support vector machine, respectively.

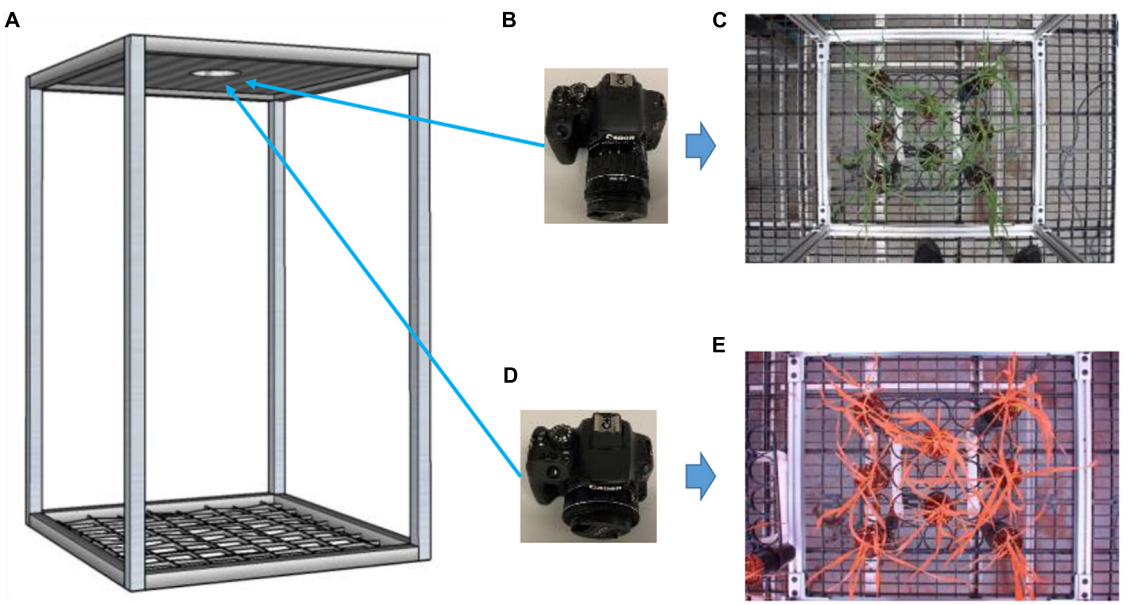


FIGURE 2 | Experimental setup for image collection. A fabricated frame (**A**; 60 × 60 × 90 cm) to hold cameras on the top at the same locations and hold rack of pots at the bottom; a color camera **(B)** with a sample collected image **(C)**; and a color-infrared camera **(D)** with a sample collected image **(E)**.

images. Manually processing the images to create the paired dataset presents many issues: (i) manually cropping both the color image and the CIR image can be a laborious process; (ii) users' manual switching between the two images is inefficient; and (iii) manual cropping method is inaccurate since it is very difficult to replicate the same ROI onto the corresponding image.

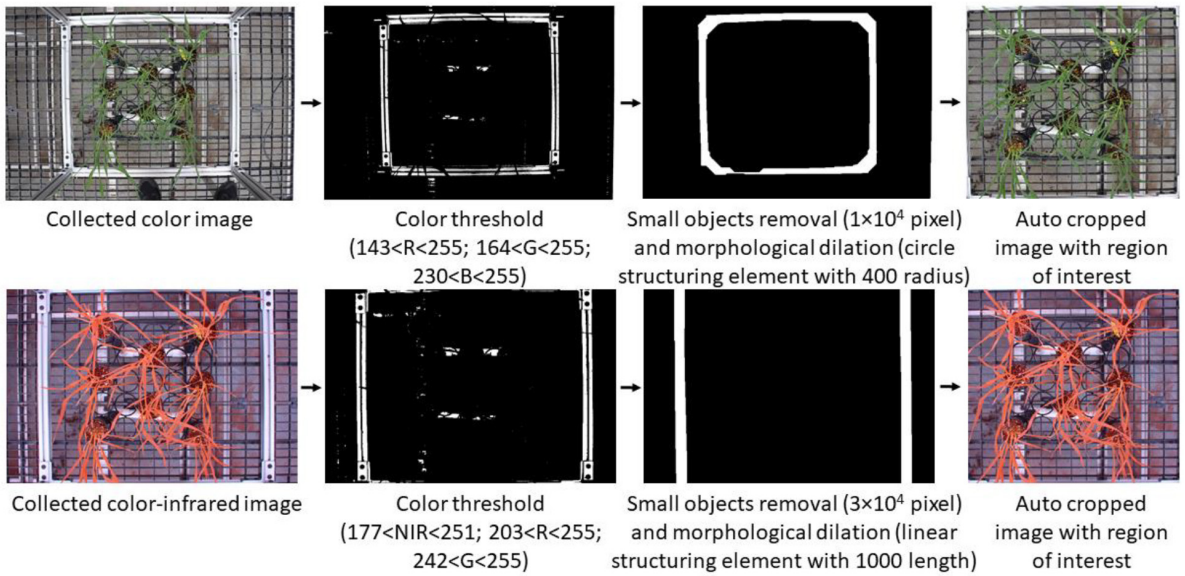


FIGURE 3 | Automatic procedures of using image process techniques to generate the corresponding two types of images with the same view. NIR, near-infrared; R, red; G, green; B, blue.

To address these issues, we developed a webtool that can expedite the workflow and improve the process accuracy. The graphical user interface (GUI) of the developed webtool is shown in **Figure 4A** and can be accessed *via* this webpage.[1] Following the GUI instructions, users needed to upload a pair of images (auto-cropped color and CIR image; **Figure 3**). The webtool would then automatically resize the two images to make their dimensions similar (image size; **Figure 4B**). Then, users can draw the ROI (any closed polygonal or irregular shape) at the top image (red irregular shape in **Figure 4C**), after which the corresponding image of the other type would be generated and saved (a sample pair shown in **Figure 4D**). This free webtool can be accessed by users for a similar image processing workflow.

Visual Disease Classification

After the image dataset for each type of disease was generated, each pair of the image was visually classified into one of the three following classes: control (disease-free), disease light (light infection), and disease severe (severe infection). The standards used for classifying leaf rust disease grade are shown in **Supplementary Appendix I**: If no disease symbol was shown, it was classified as control; if the rust severity level was below 10 of the *modified Cobb Scale* B, it was classified as disease light; otherwise, it was classified as disease severe (Peterson et al., 1948; Gebremariam et al., 2016). Samples of visually graded leaf rust diseases with different severity levels are shown in **Supplementary Appendix I**. For the tan spot visual grading, the following protocols were followed: If no disease symbol was shown, it was classified as control; if the disease area (discolored portion) was less than 30% of the total leaf area, it was classified as disease light; otherwise, it was classified as disease severe. Samples of visually graded tan spot diseases with different severity levels are shown in **Supplementary Appendix II**. Since visual disease classification requires domain knowledge, in this study, three individual graders were trained by professional plant pathologists and then voted for the classification of each image. For each image, the grade with more than two votes was assigned as the final grade. There were no cases that the three graders assigned three different grades for the same image.

Segmentation of Color and Color-Infrared Images

After creating and grading the datasets, the segmentation of the plant from the noisy background, including fertilizer, peat, plastic grid, and aluminum frame, was performed. Color images were first converted to Lab (L for lightness and a and b for the color dimensions) format, and then, proper thresholding was applied to generate a binary image. After removing small area noises, a binary mask was generated, which was applied to the original image to obtain the segmented color image (**Figure 5**).

Since the paired color and CIR images are of the same size, the easiest approach to segment the CIR image was to directly apply the binary mask generated during the color image segmentation to the corresponding CIR image. However, considering that most studies used the raw CIR images for disease detection (Bajwa and Tian, 2002; Yang, 2020) and few studies reported CIR image segmentation methods, our interest was to develop a general approach for CIR image segmentation that could be referred by other researchers. After preliminarily testing several approaches, the K-means clustering algorithm was selected. A key parameter while applying this algorithm was the selection of proper number of clusters. In this study, "3" was applied as pixels can be categorized into three clusters, namely, plant, background, and noise (**Figure 5**). Since three images (clusters) were randomly

[1]https://github.com/jithin8mathew/RGB_CIR_imageCropping_tool

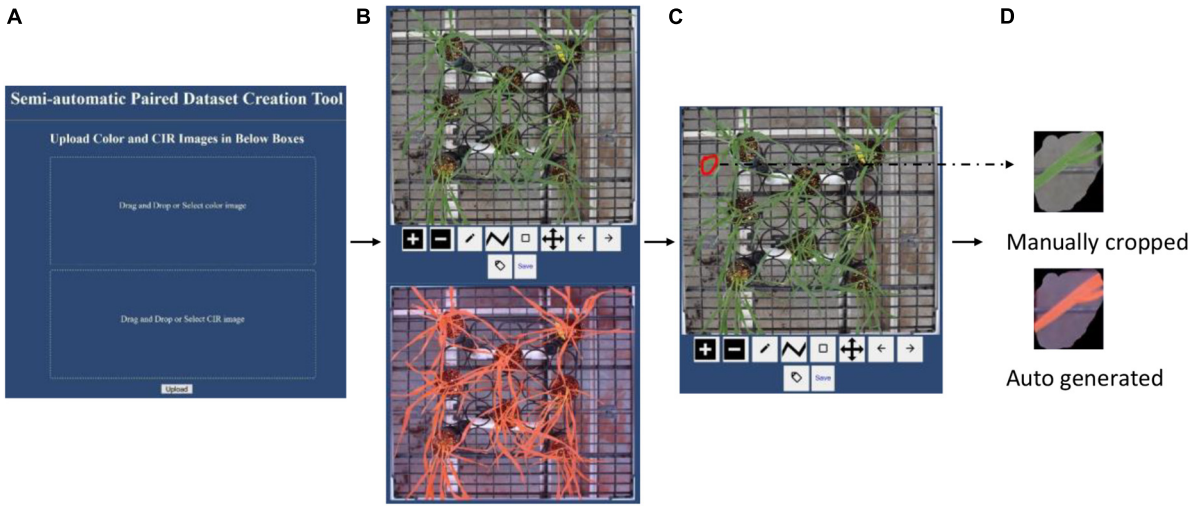


FIGURE 4 | Introduction to the webtool developed to expedite the paired dataset generation accurately. **(A)** Graphical user interface, **(B)** uploaded a pair of images with different formats (color and color-infrared), **(C)** manual image cropping on color image (red irregular shape), and **(D)** sample of a paired dataset.

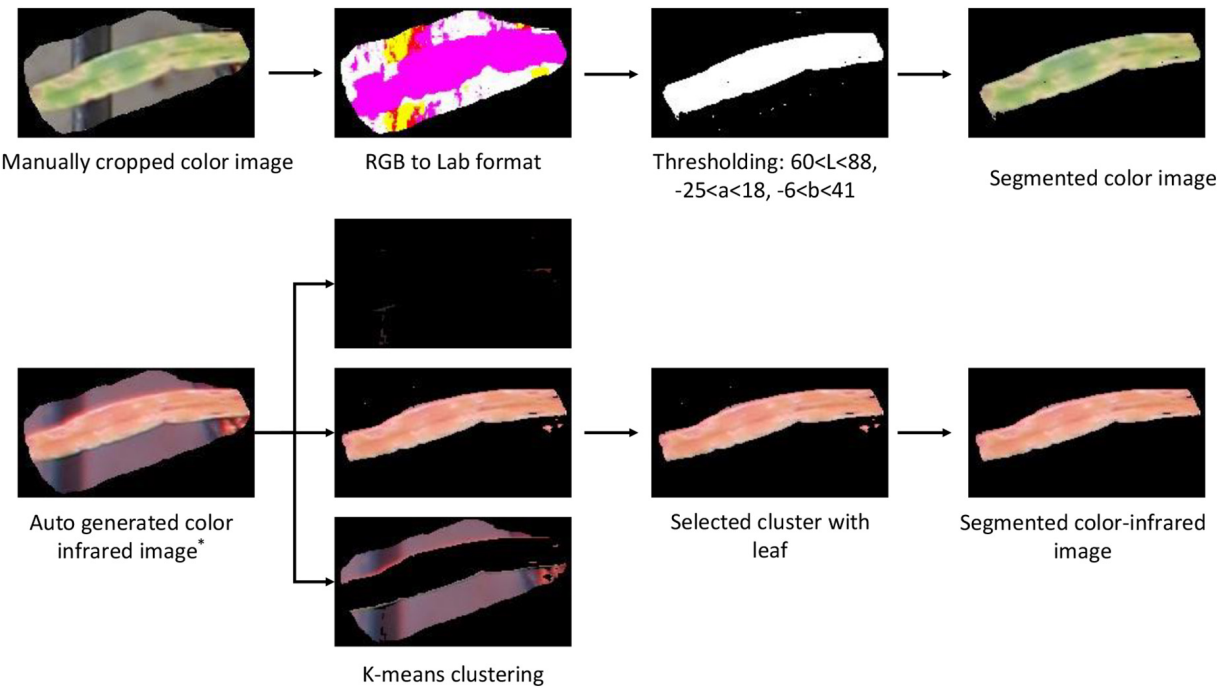


FIGURE 5 | Image segmentation procedures of the color and color-infrared paired images. * refers to the auto-generated image from the webtool (**Figure 4**).

generated after implementing the algorithm, it was necessary to develop a solution that would automatically select the proper cluster with plants (not background or noise). In the NIR channel, plants had a stronger signal over the background and noise, which made it a good parameter to differentiate the plant cluster from noise and background clusters. The average intensity for the NIR channel of each image was calculated, and the image with the highest value was selected as the plant cluster. After small objects as noises were removed, the leaf in the CIR image was segmented, and then, the dataset was prepared for further analysis. The dataset size is shown in **Table 1**.

Handcrafted Features Extraction

Based on domain knowledge and reported results (Lu and Lu, 2018; Wang et al., 2019), color, vegetation fraction, and texture were the extracted HFs features in this study. Although both color and CIR images contain red and green channels, based on our preliminary comparisons of the same channel images from two

TABLE 1 | Datasets (region of interest) of paired images and their sizes.

Leaf rust disease dataset	Number of images	Tan spot disease dataset	Number of images
Free color	226	Free color	237
Free color-infrared	226	Free color-infrared	237
Light color	171	Light color	182
Light color-infrared	171	Light color-infrared	182
Severe color	200	Severe color	200
Severe color-infrared	200	Severe color-infrared	200

cameras, they were not exactly the same (probably due to slight differences in band center and width of the two cameras).

For the color images in RGB format, they were also converted into HSI (hue, saturation, intensity) and Lab formats, and the normalized average intensity of each channel was calculated (a total of 9 HFs). For the CIR images, since the original image consists of 3 channels (e.g., R, G, and NIR), the normalized average intensity of each channel was calculated. Then, the CIR image was converted into HSI and Lab formats, with another 6 color features obtained (a total of 9 HFs).

After extracting the color features, the vegetation fraction features were extracted. For the color image, it included Normalized Difference Index (NDI), Excess Green (E × G), Excess Red (E × R), Color Index of Vegetation Extraction (CIVE), Modified Excess Green (ME × G), and Normalized Excess Green (NE × G). For the CIR image, due to the lack of blue channel information, the features of E × G, CIVE, ME × G, and NE × G could not be extracted. However, with the NIR channel information, NDVI and Green Normalized Difference Vegetation Index (GNDVI) features were obtained. For the texture features, they were extracted using the gray-level co-occurrence matrix (GLCM), including correlation, contrast, dissimilarity, energy, entropy, and homogeneity. Details of all the features extracted, as well as their calculation formulas, have been described in Wood et al. (2012); Wang et al. (2019), and Aballa et al. (2020).

Deep Features Extraction

In contrast to HFs that require domain knowledge to decide which type of features to extract, the extraction of DFs is domain knowledge-free. Although different trained CNNs can be used to extract DFs, it was unknown which one could extract appropriate DFs that can better represent crop diseases. Hence, several DL models, including AlexNet, VGG16, ResNet101, GoogLeNet, and Xception, were used to extract DFs, which were then fed into a classifier to select the one with the highest accuracy. Since model training is time-consuming, this study took advantage of these trained models for DFs extraction (Aballa et al., 2020). Since each model has many deep layers (consisting of CNNs and fully connected layers), it was necessary to decide which layer to use for DFs extraction. Since previous studies demonstrated that shallow layers mainly reserved spatial and general shape information (might not be significantly related to disease detection) (Jiang and Li, 2020), this study extracted DFs using the layer before the last fully connected layer of each model. The layer name and the number of extracted DFs for each model are presented in **Table 2**.

TABLE 2 | Information of deep learning models used for extracting deep features.

Model information	AlexNet	VGG16	ResNet101	GoogLeNet	Xception
Number of deep layers	8	16	101	22	71
Feature pooling layer name	drop7	drop7	pool5	pool5-drop_7 × 7_s1	avg_pool
Number of features extracted	4,096	4,096	2,048	1,024	2,048

Parallel Feature Fusion

Many data fusion techniques have been used to improve model accuracy. One approach is to first register images from different sources and then fuse them using the Laplacian Pyramid Transform (LPT) or Fuzzy Logic into one composite image (Bulanon et al., 2009). Then, the extracted features from the composite images are fed into classifiers. In another approach, on a first step decisions based on different feature sets (from different types of images) are made, after which the decisions are reconciled or combined to generate a global decision (decision fusion) (Peli et al., 1999; Yang et al., 2003). In yet another approach, features from different types of images are fused parallelly (concatenated) and then fed into classifiers (Yang et al., 2003; Khan et al., 2018). In our case, preliminary tests revealed that the first method resulted in a poor performance, which was probably caused by the loss of color information during the fusion process. Since the second approach has not been extensively used and the performance is unavailable, we decided to use the third method for its robustness and proved performance, which included all the information from both types of images (Khan et al., 2018).

Feature Selection and Classifier

The relevance of extracted features (HFs or DFs) for the classification is unknown beforehand. Feeding irrelevant features to the model would decrease the accuracy, as well as increase the computation load. To select relevant features, ReliefF algorithm was applied to calculate the weights of individual features (Kononenko et al., 1997). The ReliefF algorithm uses K nearest neighbors (KNN) for the weight calculation (the study used $k = 3$). The higher the weight of a feature, the more relevant it is to the classification. Since a negative weight indicates an insignificant role, this study used only features with positive weights for modeling.

A large number of classifiers have been used in addressing classification issues, including SVM, neural network (NN), random forest (RF), and KNN. Among these classifiers, an SVM classifier (multi-class) (Aballa et al., 2020; Zhang et al., 2020a) was selected because many studies have shown that it outperformed others (Sajedi et al., 2019; Jahan et al., 2020). For the diagnosis of different types of diseases (e.g., leaf rust, tan spot, and leaf

rust + tan spot), the dataset was randomly partitioned into training (80%) and testing (20%) for model development.

In this study, Python language (V3.8) was used to develop the semiautomatic webtool (see text footnote 1) to assist paired image dataset creation (**Figure 4**). For all other data processing (e.g., image segmentation, feature extraction, and model development and execution), MATLAB® 2019a (The Mathworks, Inc., Natick, MA, United States) was used. A desktop computer was used for data processing, which was configured with Windows 10 OS, Intel(R) Core(TM) i7-8700 CPU, 32 GB RAM, Intel(R) UHD Graphics 630, and 16 G GPU memory.

RESULTS AND DISCUSSION

Diagnosis Accuracies Based on Handcrafted Features

Disease detection accuracies based on HFs from different types of images (color and CIR) and their parallel fusion are shown in **Figure 6**. For the statistical analysis and visualization of the whole data, the effects of features from different types of images for the same disease (**Figure 6A**) and the effects of disease type for the features extracted from a certain type of image (**Figure 6B**) are presented. For leaf rust and leaf rust + tan spot (**Figure 6A**) disease diagnosis, the CIR (accuracy about 53%) did not perform as satisfactorily as the color images (accuracy about 60%). One possible reason was that the leaf rust disease's symptoms were relatively small in size, and the difference between the diseased and healthy regions was unobvious of the CIR images. This assumption is supported by the results regarding tan spot disease, where the performance of CIR and color images was not significantly different (**Figure 6A**), as the tan spot symptoms were large and more obvious, hence identified with a better accuracy (about 75%). Compared with the single-source features from CIR or color images, for all the three types of diseases, the parallelly fused features resulted in the highest accuracies. Since one type of image can only represent partial information, the parallel fused features from images collected by both cameras provided more meaningful information and features that best described the diseases. For the leaf rust, tan spot, and leaf rust + tan spot, the accuracy improvement using parallelly fused features over CIR image features was 21, 8, and 27%, respectively, and over color image features was 9, 10, and 4%, respectively. Such good accuracy improvements indicate the superiority of applying parallel feature fusion techniques for disease diagnosis.

Regarding detection accuracy for different diseases (**Figure 6B**) with the same type of features, it can be observed that tan spot was consistently the disease detected with the highest accuracy across the three types of features. The obvious difference of the tan spot symptoms from the healthy portion of the leaves and large area of discolored tissues might be the reason for the more accurate detection. On the contrary, the symptoms of leaf rust are usually small in discolored area (not significant). The overall accuracy for the leaf rust + tan spot is a little higher than that of the leaf rust alone, which is because the combined dataset contained the tan spot disease sub-dataset as well. For the CIR features, color features, and parallelly fused features, the tan spot detection accuracies were 46, 28, and 31% higher over leaf rust, respectively, and 47, 18, and 25% higher over leaf rust + tan spot, respectively.

The confusion matrices presented in **Figure 7** provide more detailed information on the classification/misclassification results. For the leaf rust disease severity detection, the model had difficulties in classifying the leaf rust light (light disease) correctly. A total of 31 cases (8, 12, and 11 from **Figures 7A–C**, respectively) of leaf rust light were misclassified as leaf rust control (disease-free), and a total of 33 cases (10, 18, and 5 from **Figures 7A–C**, respectively) of leaf rust light were miscategorized as leaf rust severe. The light condition may have played a big role on those results, making it very difficult to accurately assess the disease occurrence and severity. Our findings further demonstrated that disease detection on its early stages of development is challenging, which is supported by previous literature reports (Singh and Misra, 2017). For the tan spot disease detection, the major misclassifications occurred as light disease predicted as severe (30 cases; 12, 6, and 12 from **Figures 7D–F**, respectively). The difficulties can further support the previous assessment that it is a challenge to accurately detect and assess severity on its early stage of infection. The results (**Figures 7G–I**) showed a good performance in identifying the disease type (tan spot or leaf rust). For the CIR features, color features, and parallelly fused features, the disease misclassification rates (leaf rust classified as a tan spot or tan spot classified as leaf rust) were 19% (28 cases in **Figure 7G**), 11% (16 cases in **Figure 7H**), and 9% (14 cases in **Figure 7I**), respectively. Thus, the parallelly fused features have a more satisfactory performance in disease type identification. This piece of information is critical for researchers and farmers to select proper chemicals for disease management.

Model Selection for Deep Features Extraction

The diagnosis accuracies based on DFs from five DL models are shown in **Figure 8**. For all the 15 settings (5 DL models × 3 types of diseases), the fused features resulted in the highest accuracy in eight settings (e.g., AlexNet TS and LR + TS, GoogLeNet TS and LR + TS, ResNet101 LR + TS, VGG16 TS, and Xception TS and LR + TS, where TS and LR are tan spot and leaf rust, respectively), while for the other seven settings, fused features together with color image features lead to the highest accuracies. The results indicate that, similar to the HFs, parallelly fused deep features could increase the model accuracy over deep features from a single type of image.

To assist decision-making on which DL model should be selected for DFs extraction, the experimental results of **Figure 8** were rearranged as shown in **Figure 9** for better comparison of different models and statistical analysis. In each of the nine settings (3 diseases × 3 types of features), DFs extracted by ResNet101 consistently resulted in the highest accuracy (letter *a* in all 9 settings). The high diagnosis accuracy was due to the extracted features that were good representations of the crop diseases. A potential reason that the DFs by ResNet101 outperformed the DFs extracted by AlexNet, VGG16,

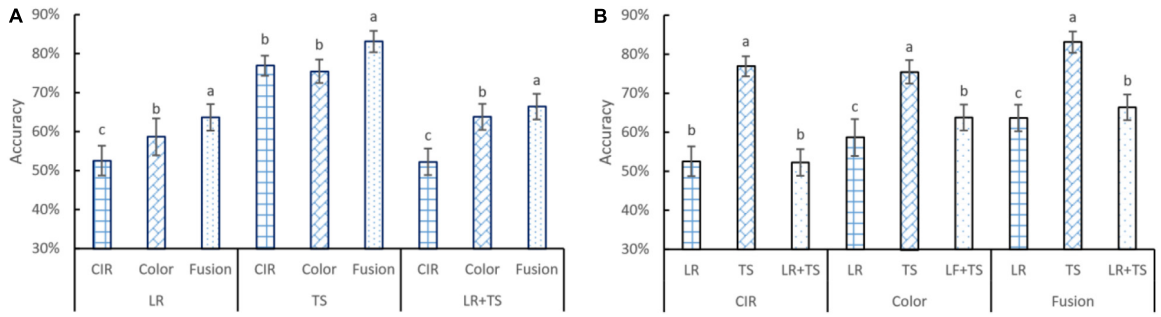


FIGURE 6 | Accuracy performance of features from color and color-infrared image, and their parallel fusion on leaf rust (LR), tan spot (TS), and LR + TS (two diseases combination) detection in terms of different image types **(A)** and diseases **(B)**. Whiskers on bars represent two standard deviations calculated from 20 replicates. Bars with different letters are significantly different by Tukey's test at a significance level of 0.05.

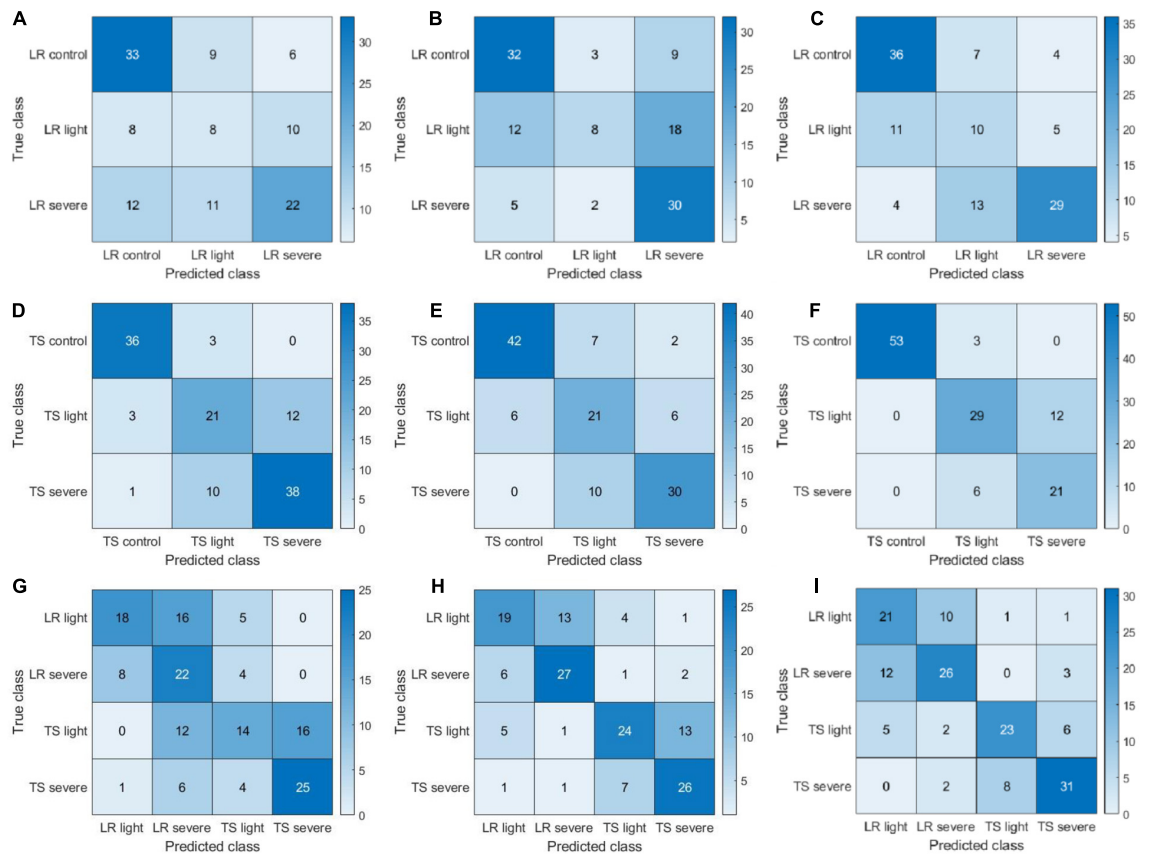


FIGURE 7 | Confusion matrices of color-infrared (CIR) image features, color image features, and parallelly fused features on leaf rust (LR), tan spot (TS), and LR + TS (their combination) disease detection based on handcrafted features. **(A–I)** Represent a subset of three confusion matrices for CIR image features, color image features, and parallelly fused features on LR, TS, and LR + TS disease detection, respectively.

GoogLeNet, and Xception was because the ResNet101 has more deep layers—ResNet101, AlexNet, VGG16, GoogLeNet, and Xception consist of 101, 8, 16, 22, and 71 deep layers, respectively (**Table 1**). With more layers, the extracted DFs could represent more detailed (fine) information of the plant diseases, while the features from shallow layers mainly reserved spatial and general information (Jiang and Li, 2020). Hence, the ResNet101 should be used for DFs extraction to serve the purpose of wheat disease diagnosis.

Since DFs extracted by ResNet101 would lead to the highest diagnosis accuracy compared with other DL models, its performance was further studied in the form of confusion matrices to reveal the detailed classification/misclassification results (**Figure 10**). For the leaf rust disease diagnosis, a majority

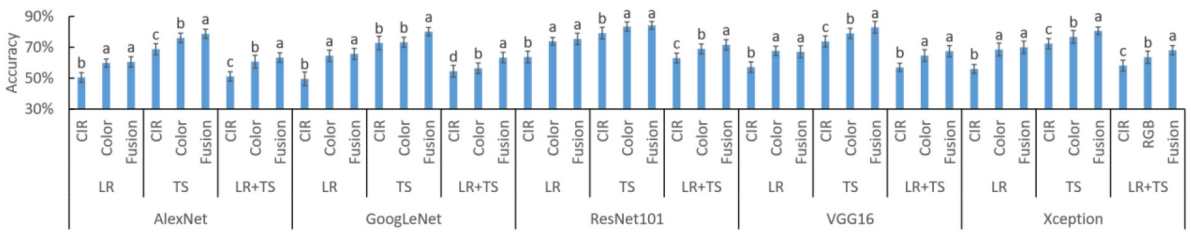


FIGURE 8 | Performance of deep features (DFs) from different types of images and their parallel fusion on different wheat diseases detection. DFs were extracted by five different models. CIR, color-infrared images; LR, leaf rust; TS, tan spot; and LR + TS, combined disease datasets (without the control datasets). Whiskers on bars represent two standard deviations calculated from 20 replicates. Bars with different letters are significantly different by Tukey's test at a 0.05 significance level.

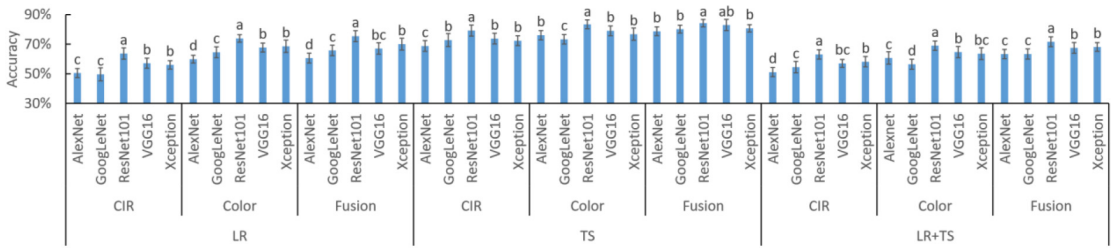


FIGURE 9 | Accuracies of deep features (DFs) extracted by different deep learning models on different wheat disease diagnoses. Fusion means the parallelly fused features of color-infrared (CIR) and color images. LR, leaf rust; TS, tan spot; and LR + TS, combined disease dataset (without the control datasets). Whiskers on bars represent two standard deviations calculated from 20 replicates. Bars with different letters are significantly different by Tukey's test at a significance level of 0.05.

of the misclassification cases occurred as the severe infection cases predicted as light (33 cases consisting of 14, 13, and 6 cases from **Figures 10A–C**, respectively). The pattern of misclassification was different from the HFs as mentioned in **Figure 7**, where light infection cases were misclassified as severe or free of infection. The different types of misclassification indicate that the DFs represented the images differently from the HFs. For the tan spot disease diagnosis, most misclassifications that happened as severe infections were predicted as light (34 cases consisting of 13, 10, and 11 from **Figures 10D–F**, respectively). This type of misclassification was also different from the HFs (**Figure 7**), supporting the previous assessment that DFs represented the images differently from the HFs. The DF had a satisfactory performance in distinguishing the combined scenario of two diseases, and the misclassification ratios were only 6.6% (10 cases in **Figure 10G**), 5.3% (8 cases in **Figure 10H**), and 4.6% (7 cases in **Figure 10I**) for the CIR image DFs, color image DFs, and the parallelly fused DFs, respectively.

Accuracy Comparison Between Handcrafted and Deep Features

Results shown so far have demonstrated that parallel feature fusion could improve the model accuracy for wheat disease diagnosis for both HFs and DFs, and the DFs extracted by ResNet101 resulted in higher accuracy over the other four models, namely, AlexNet, GoogLeNet, VGG16, and Xception. To make a better assessment of that, we did a side-by-side comparison of HFs and DFs (extracted by ResNet101) on the detection of diseases using the parallelly fused features (**Figure 11**). For the leaf rust and leaf rust + tan spot disease diagnosis, DFs resulted in higher accuracies of 19 and 8% over HFs, respectively. A potential explanation for those results is that the symptoms for leaf rust were not very obvious (could not be manually selected), and DFs were able to extract fine features that could better represent the diseases. For the tan spot disease diagnosis, the accuracies by DFs were not significantly different from HFs, which might be because the TS symptoms were clear and obvious. Overall, it is recommended to use the DFs, instead of HFs, for wheat disease detection, coupled with the parallel feature fusion technique.

The accuracies of several wheat disease detection (e.g., smut, leaf blotch, and black chaff) studies varied greatly from 50 to 99% (Lu et al., 2017; Su et al., 2018, 2019; Wiesner-Hanks et al., 2019; Abdulridha et al., 2020; Mi et al., 2020). From this study, the recommended application of parallel fusion of CIR image DFs and color image DFs extracted from ResNet 101 resulted in the accuracies of 75, 84, and 72% for wheat leaf rust, tan spot, and leaf rust + tan spot disease detection, respectively (**Figure 11**). Due to the differences in the datasets applied and models employed in this study as compared with other published studies, it is nearly impossible to make a direct and objective comparison. However, the accuracy of the methodologies by Chen et al. (2018); Su et al. (2018), and Wiesner-Hanks et al. (2019) might be improved by incorporating the outcomes of this research—DFs (from ResNet101) coupled with parallel feature fusion for diagnosis accuracy improvement.

This study developed the methodology specifically for greenhouse applications. However, it also has the potential to be applied for field use in real-time mode, which can help

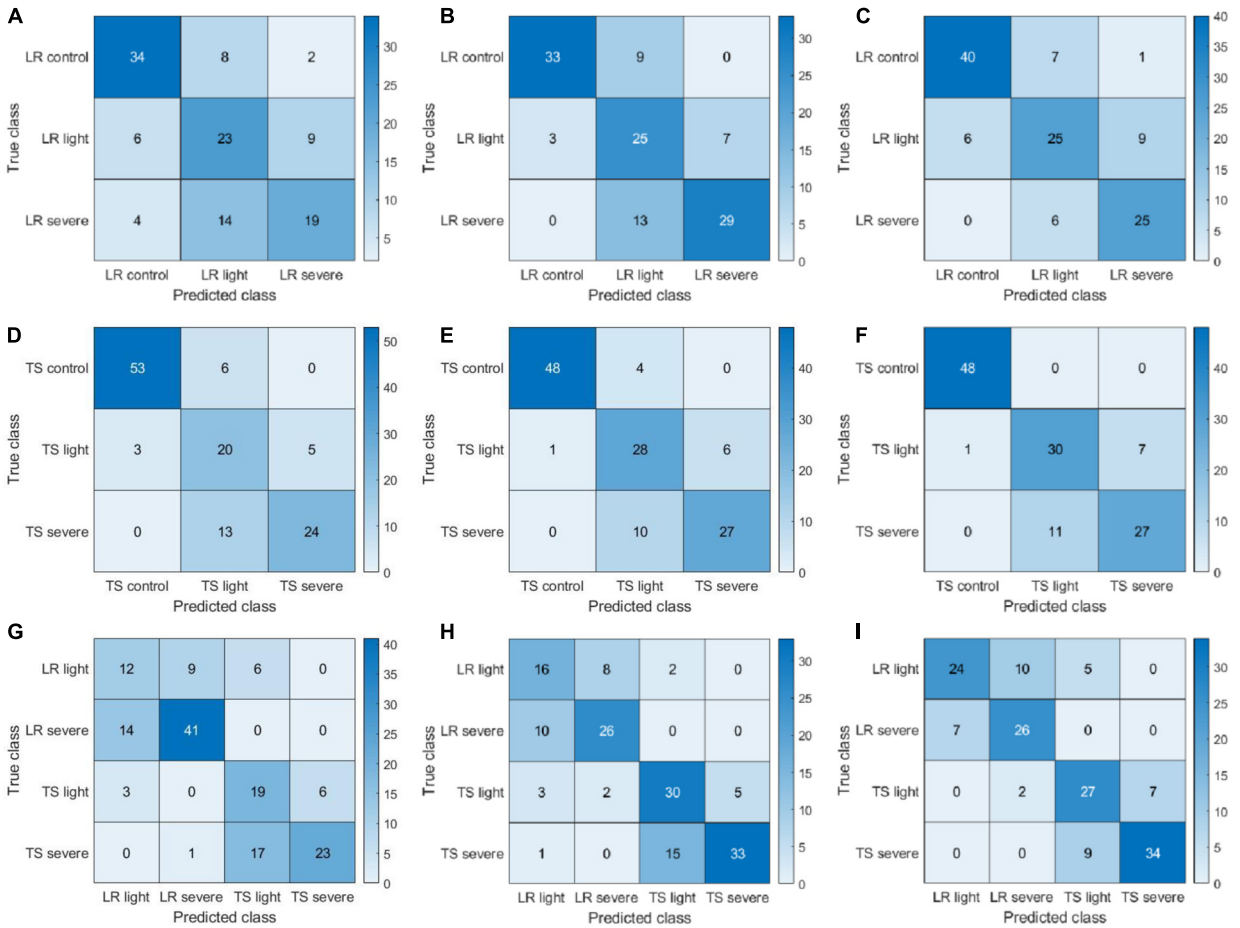


FIGURE 10 | Confusion matrices of color-infrared (CIR) image deep features (DFs), color image DFs, and their parallelly fused DFs on leaf rust (LR), tan spot (TS), and LR + TS (their combination) disease detection. DFs extracted by ResNet101 and sub-figures **(A–I)** represent three confusion matrices of CIR HFs, color HFs, and fusion for LR, TS, and LR + TS, respectively.

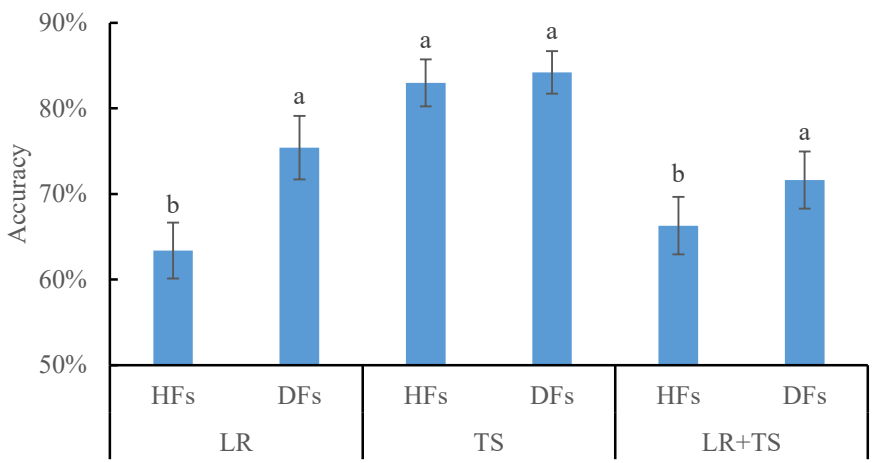


FIGURE 11 | Comparison of handcrafted features (HFs) and deep features (DFs; extracted from ResNet101) on different wheat diseases detection with parallel feature fusion. Whiskers on bars represent two standard deviations calculated from 20 replicates. Bars with different letters are significantly different by Tukey's test at a significance level of 0.05.

breeders, plant scientists, and growers to obtain the wheat disease conditions. Thus, the current methodology should be tested using field data. Considering the variable lighting conditions during infield use, a color calibration/adjustment procedure at the beginning of the data process should be added to improve the model's robustness (Sunoj et al., 2018). Furthermore, a desktop was used for the data process in this study, which should be replaced by an embedded system for infield use. Thus, a trade-off between model size, computation time, and model accuracy should be made, instead of using one parameter (accuracy) to judge the model performance.

CONCLUSION

A methodology for the diagnosis of leaf rust, tan spot, and leaf rust + tan spot diseases with handcrafted and deep features from the color image, color-infrared (CIR) image, and their parallel fusion along with SVM classifier was successfully developed and compared. A webtool was developed, hosted (see text footnote 1), and used in this study for paired datasets (the same view for color and CIR images) creation. Fused features (parallel mode in this study obtained *via* concatenating) resulted in a higher disease detection accuracy over the features from a single type image (either color or CIR). It was found that deep features (automatically selected by DL algorithms with free domain knowledge) generated higher diagnosis accuracies over handcrafted features (manually selected using domain knowledge), due to extraction of fine features by DFs that would be missed by HFs. In addition, while selecting DL models for DFs extraction, it is recommended to use the efficient ResNet101 DL model generating more deep layers, as shallow features can only reserve spatial and general information. The developed methodology based on DFs and parallel feature fusion efficiently detected wheat disease conditions with accuracies of 74, 84, and 72% for leaf rust, tan spot, and leaf rust + tan spot, respectively. This methodology, which can be readily used in greenhouse applications by plant pathologists, breeders, and other users, presents a pathway toward the development of automatic and objective wheat disease diagnosis applications. Furthermore, the field application of the methodology can be achieved with further tests of field data and fine-tuning of model parameters.

This study successfully and satisfactorily segmented color and CIR images using the developed general algorithms. However, the segmentation results between color and CIR images were not compared. Future studies, such as comparing overlapping ratio, should be conducted in this field. Furthermore, the mask generated for segmentation color images should be tested on the CIR image, and vice versa. This study took advantage of ReliefF for the feature section, and under some conditions, the elimination of features may not improve the model accuracy. Therefore, future studies should compare the model accuracy between the feature selection and non-selection. This study mainly focuses on the model accuracy, and it lacks a comprehensive comparison among different models, such as training time and model size. Future research should compare the models more comprehensively. This study took advantage of the SVM as the classifier for its proven performance. Further studies should be conducted to compare the performance of different classifiers, such as neural network and random forest.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

ZZ contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, writing original draft, supervision, reviewing, and editing. PF contributed to funding acquisition, writing, reviewing, and editing. AF, ZL, and CI contributed to resources, writing, reviewing, and editing. XH and HK contributed to writing, reviewing, and editing. NJ and SS contributed to data curation. JM contributed to software. All authors contributed to the article and approved the submitted version.

FUNDING

This work was jointly supported by the North Dakota Agricultural Experiment Station (NDAES) Precision Agriculture Graduate Research Assistantship and USDA Agricultural Research Service (Project No. 6064-21660-001-32S, Project accession No. 435589).

ACKNOWLEDGMENTS

We would like to acknowledge Kalin Mogen and Kenton Jensen for their greenhouse work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.834447/full#supplementary-material

REFERENCES

Aballa, A., Cen, H., Wan, L., Mehmood, K., and He, Y. (2020). Nutrient status diagnosis of infield oilseed rape *via* deep learning-enabled dynamic model. *IEEE Trans. Ind. Inf.* 2020:9736. doi: 10.1109/TII.2020.3009736

Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., and Roberts, P. (2020). Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precis. Agric.* 21, 955–978.

Ashourloo, D., Mobasheri, M. R., and Huete, A. (2014). Developing two spectral disease indices for detection of wheat leaf rust

(*Pucciniatriticina*). *Remote Sens.* 6, 4723–4740. doi: 10.3390/rs6064723

Azadbakht, M., Ashourloo, D., Aghighi, H., Radiom, S., and Alimohammadi, A. (2019). Wheat leaf rust detection at canopy scale under different LAI levels using machine learning techniques. *Comput. Electron. Agric.* 156, 119–128. doi: 10.1016/j.compag.2018.11.016

Bajwa, S. G., and Tian, L. (2001). Aerial CIR remote sensing for weed density mapping in a soybean field. *Trans. ASAE* 44:1965. doi: 10.13031/2013.6995

Bajwa, S. G., and Tian, L. (2002). Multispectral CIR image calibration for cloud shadow and soil background influence using intensity normalization. *Appl. Eng. Agric.* 18:627. doi: 10.13031/2013.10148

Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53. doi: 10.1016/j.compag.2018.08.013

Bolton, M. D., Kolmer, J. A., and Garvin, D. F. (2008). Wheat leaf rust caused by *Puccinia triticina*. *Mol. Plant Pathol.* 9, 563–575. doi: 10.1111/j.13643703.2008.00487.x

Bravo, C., Moshou, D., West, J., McCartney, A., and Ramon, H. (2003). Early disease detection in wheat fields using spectral reflectance. *Biosyst. Eng.* 84, 137–145. doi: 10.1016/S1537-5110(02)00269-6

Bulanon, D., Burks, T., and Alchanatis, V. (2009). Image fusion of visible and thermal images for fruit detection. *Biosyst. Eng.* 103, 12–22. doi: 10.1016/j.biosystemseng.2009.02.009

Carlson, T. N., and Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* 62, 241–252. doi: 10.1016/S0034-4257(97)00104-1

Castanedo, F. (2013). A review of data fusion techniques. *Sci. World J.* 2013, 1–19. doi: 10.1155/2013/704504

Chen, D., Shi, Y., Huang, W., Zhang, J., and Wu, K. (2018). Mapping wheat rust based on high spatial resolution satellite imagery. *Comput. Electron. Agric.* 152, 109–116. doi: 10.1016/j.compag.2018.07.002

Curtis, T., and Halford, N. (2014). Food security: The challenge of increasing wheat yield and the importance of not compromising food safety. *Ann. Appl. Biol.* 164, 354–372. doi: 10.1111/aab.12108

De Wolf, E. D. (2008). *Agricultural Experiment Station and Cooperative Extension Service*. Stillwater, Ok: OSU.

Figlan, S., Ntushelo, K., Mwadzingeni, L., Terefe, T., Tsilo, T. J., and Shimelis, H. (2020). Breeding wheat for durable leaf rust resistance in southern africa: Variability, distribution, current control strategies, challenges and future prospects. *Front. Plant Sci.* 11:549. doi: 10.3389/fpls.2020.00549

Franke, J., and Menz, G. (2007). Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precis. Agric.* 8, 161–172. doi: 10.1007/s11119007-9036-y

Gaikwad, V. P., and Musande, V. (2017). Wheat disease detection using image processing, in: 2017 1st Int. *Conf. Intell. Syst. Inf. Manage. IEEE* 2017, 110–112. doi: 10.1109/ICISIM.2017.8122158

Gebremariam, T. G., Woldeab, G., and Selvaraj, T. (2016). Distribution, physiologic races and reaction of wheat cultivars to virulent races of leaf rust (*Puccinia triticina Eriks* and *Henn.*) in south eastern zone of tigray, ethiopia. *Int. J. Life Sci.* 4, 1–21.

Huang, X., Xu, H., Wu, L., Dai, H., Yao, L., and Han, F. (2016). A data fusion detection method for fish freshness based on computer vision and near-infrared spectroscopy. *Anal. Methods* 8, 2929–2935. doi: 10.1039/C5AY03005F

Jahan, N., Flores, P., Liu, Z., Friskop, A., Mathew, J. J., and Zhang, Z. (2020). *Detecting and distinguishing wheat diseases using image processing and machine learning algorithms*, St. Joseph: American Society of Agricultural and Biological Engineers. doi: 10.13031/aim.202000372

Jiang, Y., and Li, C. (2020). Convolutional neural networks for image-based highthroughput plant phenotyping: A review. *Plant Phenomics* 2020:4152816. doi: 10.34133/2020/4152816

Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A. D., et al. (2017). Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput. Electron. Agric.* 138, 200–209. doi: 10.1016/j.compag.2017.04.013

Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., et al. (2018). CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. *Comput. Electron. Agric.* 155, 220–236. doi: 10.1016/j.compag.2018.10.013

Kolmer, J. (1996). Genetics of resistance to wheat leaf rust. *Annu. Rev. Phytopathol.* 34, 435–455. doi: 10.1146/annurev.phyto.34.1.435

Kononenko, I., Simec, E., and RobnikSikonja, M. (1997). Overcoming the myopia of inductive˙ learning algorithms with relieff. *Appl. Intell* 7, 39–55.

Lehmann, J. R. K., Nieberding, F., Prinz, T., and Knoth, C. (2015). Analysis of unmanned aerial system-based CIR images in forestry—a new perspective to monitor pest infestation levels. *Forests* 6, 594–612. doi: 10.3390/f6030594

Lu, J., Hu, J., Zhao, G., Mei, F., and Zhang, C. (2017). An in-field automatic wheat disease diagnosis system. *Comput. Electron. Agric.* 142, 369–379. doi: 10.1016/j.compag.2017.09.012

Lu, Y., and Lu, R. (2018). Detection of surface and subsurface defects of apples using structuredillumination reflectance imaging with machine learning algorithms. *Trans. ASABE* 61, 1831–1842. doi: 10.13031/trans.12930

Luvisi, A., Ampatzidis, Y. G., and De Bellis, L. (2016). Plant pathology and information technology: Opportunity for management of disease outbreak and applications in regulation frameworks. *Sustainability* 8:831. doi: 10.3390/su8080831

McMullen, M., and Adhikari, T. (2009). *Fungal Leaf Spot Diseases of Wheat: Tan spot, Septoria/Stagonospora nodorum blotch and Septoria tritici blotch*. Fargo: North Dakota State University.

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.* 11:558126. doi: 10.3389/fpls.2020.558126

Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., and Ramon, H. (2004). Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Comput. Electron. Agric.* 44, 173–188. doi: 10.1016/j.compag.2004.04.003

Patil, J. K., and Kumar, R. (2011). Color feature extraction of tomato leaf diseases. *Int. J. Eng. Trends Technol.* 2, 72–74.

Peli, T., Young, M., Knox, R., Ellis, K. K., and Bennett, F. (1999). *Feature-Level Sensor Fusion*. Bellingham: International Society for Optics and Photonics. 332–339.

Peterson, R. F., Campbell, A., and Hannah, A. (1948). A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. *Can. J. Res.* 26, 496–500. doi: 10.1139/cjr48c-033

Qiu, R., Yang, C., Moghimi, A., Zhang, M., Steffenson, B. J., and Hirsch, C. D. (2019). Detection of fusarium head blight in wheat using a deep neural network and color imaging. *Remote Sens.* 11:2658. doi: 10.3390/rs11222658

Ransom, J. K., and McMullen, M. V. (2008). Yield and disease control on hard winter wheat cultivars with foliar fungicides. *Agron. J.* 100, 1130–1137.

Roberts, D., Smith, M., and Adams, J. (1993). Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sens. Environ.* 44, 255–269. doi: 10.1016/0034-4257(93)90020-X

Sajedi, H., Mohammadipanah, F., and Pashaei, A. (2019). Automated identification of myxobacterial genera using convolutional neural network. *Sci. Rep.* 9, 1–15. . https://doi.org/10.1038/s41598-019-54341-5., doi: 10.1038/s41598-019-54341-5

Saleem, M. H., Potgieter, J., and Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants* 8:468. doi: 10.3390/plants8110468

Salgado, J. D., Roche, E., and Paul, P. A. (2016). *Rust diseases of wheat*. Columbus, OH: Ohio State University Extension.

Sethy, P. K., Barpanda, N. K., Rath, A. K., and Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* 175:105527. doi: 10.1016/j.compag.2020.105527

Sharma, R. C., Nazari, K., Amanov, A., Ziyaev, Z., and Jalilov, A. U. (2016). Reduction of winter wheat yield losses caused by stripe rust through fungicide management. *J. Phytopathol.* 164, 671–677. doi: 10.1111/jph.12490

Shewry, P. R., and Hey, S. J. (2015). The contribution of wheat to human diet and health. *Food Energy Secur.* 4, 178–202. doi: 10.1002/fes3.64

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

Su, J., Liu, C., Coombes, M., Hu, X., Wang, C., Xu, X., et al. (2018). Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. *Comput. Electron. Agric.* 155, 157–166. doi: 10.1016/j.compag.2018.10.017

Su, J., Liu, C., Hu, X., Xu, X., Guo, L., and Chen, W. H. (2019). Spatio-temporal monitoring of wheat yellow rust using UAV multispectral imagery. *Comput. Electron. Agric.* 167:105035. doi: 10.1016/j.compag.2019.105035

Su, J., Yi, D., Su, B., Mi, Z., Liu, C., Hu, X., et al. (2020). Aerial visual perception in smart farming: Field study of wheat yellow rust monitoring. *IEEE Trans. Ind. Inf.* 2020:237. doi: 10.1109/TII.2020.2979237

Sun, Y., Jiang, Z., Zhang, L., Dong, W., and Rao, Y. (2019). SLIC-SVM based leaf diseases saliency map extraction of tea plant. *Comput. Electron. Agric.* 157, 102–109. doi: 10.1016/j.compag.2018.12.042

Sunoj, S., Igathinathane, C., Saliendra, N., Hendrickson, J., and Archer, D. (2018). Color calibration of digital images for agriculture and other applications. *ISPRS J. Photogramm. Remote Sens.* 146, 221–234. doi: 10.1016/j.isprsjprs.2018.09.015

Wan, L., Cen, H., Zhu, J., Zhang, J., Du, X., and He, Y. (2020). Using fusion of texture features and vegetation indices from water concentration in rice crop to UAV remote sensing monitor. *Smart Agric.* 2:58.

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

Wiesner-Hanks, T., Wu, H., Stewart, E., DeChant, C., Kaczmar, N., Lipson, H., et al. (2019). Millimeter-level plant disease detection from aerial photographs *via* deep learning and crowdsourced data. *Front. Plant Sci.* 10:1550. doi: 10.3389/fpls.2019.01550

Wood, E. M., Pidgeon, A. M., Radeloff, V. C., and Keuler, N. S. (2012). Image texture as a remotely sensed measure of vegetation structure. *Remote Sens. Environ.* 121, 516–526. doi: 10.1016/j.rse.2012.01.003

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, J., Yang, J., Zhang, D., and Lu, J. F. (2003). Feature fusion: Parallel strategy vs. serial strategy. *Patt. Recognit.* 36, 1369–1381. doi: 10.1016/S00313203(02)00262-5

Zhang, Z., Flores, P., Igathinathane, C., Naik, L., Kiran, R., and Ransom, J. K. (2020a). Wheat lodging detection from UAS imagery using machine learning algorithms. *Remote Sens.* 12:1838. doi: 10.3390/rs12111838

Zhang, Z., Heinemann, P. H., Liu, J., Baugher, T. A., and Schupp, J. R. (2016). The development of mechanical apple harvesting technology: A review. *Trans. ASABE* 59, 1165–1180. doi: 10.13031/trans.59.11737

Zhang, Z., Igathinathane, C., Li, J., Cen, H., Lu, Y., and Flores, P. (2020b). Technology progress in mechanical harvest of fresh market apples. *Comput. Electron. Agric.* 175:105606. doi: 10.1016/j.compag.2020.105606

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 March 2022
doi: 10.3389/fpls.2022.839269

Precision Detection of Dense Plums in Orchards Using the Improved YOLOv4 Model

Lele Wang[1,3], Yingjie Zhao[1,3], Shengbo Liu[1,3], Yuanhong Li[1,3], Shengde Chen[1,2,3] and Yubin Lan[1,2,3,4]**

[1] College of Electronic Engineering, College of Artificial Intelligence, South China Agricultural University, Guangzhou, China, [2] Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China, [3] National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology, Guangzhou, China, [4] Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX, United States

The precision detection of dense small targets in orchards is critical for the visual perception of agricultural picking robots. At present, the visual detection algorithms for plums still have a poor recognition effect due to the characteristics of small plum shapes and dense growth. Thus, this paper proposed a lightweight model based on the improved You Only Look Once version 4 (YOLOv4) to detect dense plums in orchards. First, we employed a data augmentation method based on category balance to alleviate the imbalance in the number of plums of different maturity levels and insufficient data quantity. Second, we abandoned Center and Scale Prediction Darknet53 (CSPDarknet53) and chose a lighter MobilenetV3 on selecting backbone feature extraction networks. In the feature fusion stage, we used depthwise separable convolution (DSC) instead of standard convolution to achieve the purpose of reducing model parameters. To solve the insufficient feature extraction problem of dense targets, this model achieved fine-grained detection by introducing a 152 × 152 feature layer. The Focal loss and complete intersection over union (CIOU) loss were joined to balance the contribution of hard-to-classify and easy-to-classify samples to the total loss. Then, the improved model was trained through transfer learning at different stages. Finally, several groups of detection experiments were designed to evaluate the performance of the improved model. The results showed that the improved YOLOv4 model had the best mean average precision (mAP) performance than YOLOv4, YOLOv4-tiny, and MobileNet-Single Shot Multibox Detector (MobileNet-SSD). Compared with some results from the YOLOv4 model, the model size of the improved model is compressed by 77.85%, the parameters are only 17.92% of the original model parameters, and the detection speed is accelerated by 112%. In addition, the influence of the automatic data balance algorithm on the accuracy of the model and the detection effect of the improved model under different illumination angles, different intensity levels, and different types of occlusions were discussed in this paper. It is indicated that the improved detection model has strong robustness and high accuracy under the real natural environment, which can provide data reference for the subsequent orchard yield estimation and engineering applications of robot picking work.

Keywords: object detection, YOLOv4, MobileNetV3, data balance, plum

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Yunchao Tang,
Guangxi University, China
Yusheng Wang,
Hunan Agricultural University, China
Jasper Brown,
Faculty of Engineering, The University of Sydney, Australia

****Correspondence:***
Shengde Chen
shengde-chen@scau.edu.cn
Yubin Lan
ylan@scau.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *19 December 2021*
Accepted: *09 February 2022*
Published: *11 March 2022*

Citation:
Wang L, Zhao Y, Liu S, Li Y, Chen S and Lan Y (2022) Precision Detection of Dense Plums in Orchards Using the Improved YOLOv4 Model. Front. Plant Sci. 13:839269. doi: 10.3389/fpls.2022.839269

INTRODUCTION

Plum is a characteristic fruit in South China. Its fruit is small, densely distributed, and easily blocked by other plums or branches and leaves. Plum maturity identification and picking tasks are completed manually in the current plum orchards. At present, labor costs have unprecedentedly increased, and the proportion of labor costs in total costs is also increasing, with the increase reaching up to 12–15% in 2019 (Fu et al., 2020a). In precision agriculture, labor shortage and aging labor have posed barriers to the development of the fruit industry. Considering the above, mechanized and intelligent intensive plum picking is an indispensable part of the development of the whole fruit industry.

In recent years, relevant scholars have carried out a series of research on recognizing and detecting fruits, such as apples and citrus in precision orchards (Liao et al., 2017; Wajid et al., 2018; Gurubelli et al., 2019; Mo et al., 2021). Lin G. et al. (2020) adopted partial shape matching and probabilistic Hough transform to detect fruits in the natural environment. Fu et al. (2019) achieved the fine detection of bananas by combining color, texture features, and Support Vector Machine classifier. He et al. (2020) put forward a green citrus detection method based on the deep boundary box regression forest by fusing multiscale features of color, shape, and texture. Zhao et al. (2016) combined AdaBoost classifier and color analysis to detect tomatoes in the greenhouse scene. In summary, these studies discussed previously mainly combined the traditional image processing methods and the basic characteristics of fruit color and texture. However, the data processing required a comprehensive analysis of multiple features, complex processing procedures, and poor real-time detection, which were difficult to meet the requirements of orchard information management and robotic picking.

With the rapid development of machine learning, the deep convolutional neural network (CNN) has shown excellent performance in fruits detection. Its high extraction of high-dimensional targets features makes it possible to recognize in complex environments. There are two-stage detection methods, such as Fast RCNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2016). These target detection models based on the region suggestion method adopt the final layer of the CNN to predict. Xiong et al. (2018) employed the Faster R-CNN method to detect green citrus under different illumination and sizes, and the accuracy rate reached 77.45%. Zhang et al. (2020) developed three apple recognition algorithms based on Faster R-CNN, with mean average precision (mAP) of up to 82.4%. Fu et al. (2020b) established an algorithm that is composed of ZFNet and VGG16 of Faster R-CNN architecture to detect apples in dense-leaf fruit wall trees, and the results showed that the removal of background trees with a depth filter improved fruit detection accuracy by 2.5%.

In addition, single-stage target detection methods, such as SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017, 2018), have been widely used because of their high accuracy and detection efficiency. Xue et al. (2018) adopted the YOLOv2 network to identify immature mango, which improved the detection rate while maintaining accuracy and generalization capability. Some researchers (Liu and Wang, 2020; Wang and Liu, 2021a,b) proposed the improved network models of YOLOv3 to detect the diseases and pests of greenhouse tomatoes. The proposed detection algorithm had strong robustness and high accuracy in complex orchard scenes. Tian et al. (2019) designed an improved YOLOv3 model to detect apple at different growth stages in the orchard. Kuznetsova et al. (2020) proposed a YOLOv3 apple detection algorithm with special pre-processing and post-processing. Li et al. (2020) employed the MobileNet-YOLOv3 model to detect dragon fruit in the orchard. Wu et al. (2021) proposed an improved YOLOv3 model based on clustering optimization. Liu et al. (2021) proposed a YOLOv3-SE improved method for winter jujube fruit recognition under natural environment. The mAP of the improved model increased by 2.38∼4.81% through the analysis of detection effects under different lighting conditions, occlusion, and maturity. Ji et al. (2021) proposed an apple detection method based on the improved YOLOv4, which could accurately locate and detect apples in various complex environments. Although the YOLO series networks have shown excellent performance in fruit recognition, it is difficult to detect small targets in deep feature maps due to the loss of spatial and detailed feature information. Due to the large number of model parameters, it is a very challenging task to deploy on the devices with limited resources and achieve the goal of real-time reasoning.

Compared with apple, citrus, mango, and other fruits, plum trees are mostly planted on hillsides, and their fruit growth environment is full of complexity and uncertainty. In modern precision orchards, it is more difficult to detect small targets owing to the presence of complex noise disturbance, such as changing illumination and branch and leaf occlusion. In addition, the cluster growth of the plum itself and the mixing of different maturity lead to the poor performance of existing algorithms in plum detection (Gao X. et al., 2021). Jang et al. (2021) tried to use 3D images and MATLAB R2018a to detect plums and size estimation, and this method achieved an average recognition rate of 61.9%. Pourdarbani et al. (2019) established different classifiers and majority voting rules to compare the effects of 12 different light intensities on plum images segmentation in the natural environment, and the experimental results showed that the correct classification results of the majority voting method excluding LDA were better than those of the composition method. Brown and Sukkarieh (2021) presented two datasets gathered during a robotic harvesting trial on 2D trellis plums and used them to benchmark on the four deep learning object detection architectures. Although many researchers have done extensive work on the detection of plums, the accuracy and robustness in different scenes still need to be further improved. So far, no study has been conducted on deep learning methods

Abbreviations: YOLO, you only look once; CSP, center and scale prediction; DSC, depthwise separable convolution; PWC, pointwise convolution; PANet, path aggregation network; SPP, spatial pyramid pooling; FPN, feature pyramid network; F1, the harmonic mean of the precision and recall; AP, average precision of A category; mAP, average precision of multiple categories; IOU, intersection over union; CIOU loss, complete intersection over union loss; FIOU loss, focal IOU loss; FPS, frame per second; SSD, single shot multibox detector; MobileNet-SSD, MobileNet-single shot multibox detector; UAV, unmanned aerial vehicle.

to detect dense plums in natural environments. The resources that fruit-picking robots can use in the orchard are limited. Therefore, it is necessary to explore an efficient and accurate plum recognition algorithm according to actual needs.

Aiming at the growth characteristics of plum fruit, this paper took advantage of the YOLOv4 network in target detection and combined it with the MobileNetV3 lightweight network. In the feature fusion structure, deep separable convolution was introduced to replace standard convolution, and a new convolution layer was introduced to increase the recognition performance of the model for dense small targets. Meanwhile, the Focal loss function was added to balance the contribution of different samples to the total loss. The proposed method is compared and evaluated with the other three target detection networks in different scenes to provide a reference for the yield estimation of plum and the rapid recognition of picking robots.

MATERIALS AND METHODS

Materials

Image and Data Acquisition

The experimental collection site is located in a plum orchard (23.55N, 113.59E) in Conghua District, Guangzhou City, Guangdong Province, China. The geographical location of the image acquisition is shown in **Figure 1**. The sampling device in this study is a high-resolution smartphone with a camera parameter of 40 million pixels, the exposure parameter is automatic, and the objective focus system is set to autofocus mode.

The sampling objects were plums. To collect as much information about plums in the natural environment as possible, the experimenter simulated the image capture module of the picking robot, and the handheld collection device continuously changed the shooting angle and shooting distance, hoping to collect RGB images of different colors, postures, sizes, backgrounds, and density. The experimental samples were obtained in two batches. The photographs were taken on April 24, 2021, which was a sunny day. The weather changed from light rain to cloudy from May 3 to 4, 2021. The plums were in the middle of maturity during these sessions. Most mature plums' color is red, and some immature plums' color is cyan. In total, 1,890 original images were collected under different scenes. Mature and immature plums were included in the photographs. The overall quality of the image could meet the requirements of target detection by making a visual quality assessment on the collected image data.

Dataset Production

The collected plum images have 3, 968 $\times$ 2, 976 pixels. However, the high pixel will prolong the training and processing time. This study adopted a bicubic scaling algorithm to scale image pixels to 1, 920 $\times$ 1, 440.

The Label Img, an image annotation tool, was used for manual annotation to obtain the ground truth for subsequent training. As shown in **Figure 2**, the wholly exposed plums are marked by cutting the outer part to the inside of the rectangular frame. For occluded or conglutinated plums, only the exposed parts of the image are marked. The unmarked processing was performed when the part of the image boundary or the degree of occluded plums was less than 10%. The annotation information was saved in the format of the PASCAL VOC dataset. The maturity was manually judged and marked as two types of plums, mature (plum) and immature (raw_plum).

For the marked 1,890 plum images, the original dataset was divided into the training set, validation set, and test set, where the ratio of training set to test set is 8:2. The validation set is randomly selected from 10% of the training set and does not participate in training. The training set was divided into three sub-datasets according to different collection times. Among them, sub-dataset 1 was composed of 368 image data collected on April 24, 2021, sub-dataset 2 was composed of 400 images collected on May 2, 2021, and sub-dataset 3 was composed of 744 image data collected on May 3, 2021. **Table 1** shows the data before data balance.

According to the number of plums in 1,890 labeled images, there are 10,441 mature and 4,754 immature plums labels. The proportion between the two is close to 2.2:1. It can be found

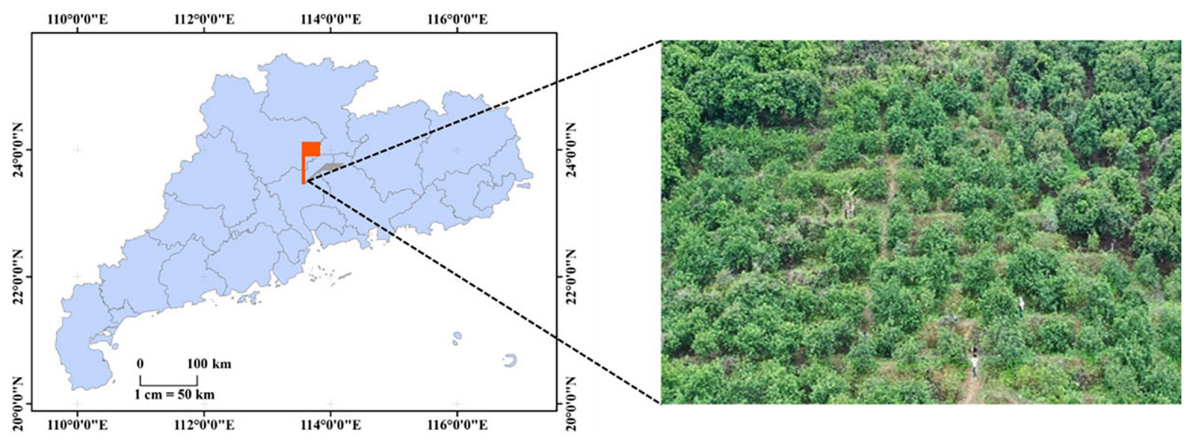


FIGURE 1 | Location of images acquisition site.

that there is a larger data imbalance between the number of mature plums and immature plums. If the network model is trained directly, it will have poor recognition performance for immature plums, resulting in the degradation of model detection ability. Therefore, it is necessary to take some measures to balance the dataset to improve the recognition ability of the model for immature plums.

Data Augmentation Method Based on Category Balance

Automatic Data Balancing Method Based on Category

Aiming at the imbalance mentioned above, this paper proposed an automatic data balancing method based on category to optimize the dataset so that the number of categories before the network model training is the same as possible (Gao J. et al., 2021). This method needs to obtain the quantitative values of all categories first, compare and select the category with the largest amount of data, and then sequentially expand the quantitative values of other categories to approach the largest category. The specific steps are as follows:

i. Suppose there is a dataset $S = [M_1, M_2, , M_i]\,[N_1, N_2, ..., N_j]^T$, where M_i denotes the number of types of samples in the dataset, and N_j denotes the number of samples in each category;
ii. Compare the sample quantity values of all categories in the dataset M_iN_j and find the maximum value M_iN_{jmax};
iii. Use M_iN_{jmax} to divide by the sample quantity value M_iN_j of the remaining category in turn, and then division C is obtained. The calculation is given in Equation 1:

$$C = \frac{M_iN_{jmax}}{M_iN_j} = [\{c_1, c_2, ..., c_{i-1}\}] \quad (1)$$

iv. Choose a data quantity expansion method, and the data quantity of residual categories become large according to division C so that the number of samples of all categories is expanded to the maximum value, and $M_iN^{'}{}_j$ is obtained, and finally, the quantity proportion of each category is close to 1;
v. The final output is the expanded dataset $T = [M_1, M_2, , M_i]\,[N^{'}_1, N^{'}_2, , N^{'}_{jmax}]T$.

According to the automatic data balancing algorithm, the number of mature groups is divided by the number of immature groups in the whole dataset, and the remainder is rounded down to get 2. Since there are different proportions of mature plum and immature plum labels in each sub-dataset, it is necessary to balance the whole sub-dataset in data balancing. Therefore, only one data amplification of sub-dataset 1 can ensure that the overall proportion of immature and mature plums in the dataset is close to 1.

Data Augmentation

To prevent overfitting or non-convergence phenomenon caused by too little training data, this study randomly combines common data augmentation methods and performs data augmentation processing on the train set, such as Gaussian blur, random rotation, random cutting off part of the image, histogram equalization, random brightness adjustment, salt, and pepper noise (Huang et al., 2020; Wu et al., 2020). The dataset is enhanced five times through the multiple random combinations of the above methods. The enhanced dataset is shown in **Table 1**. At the same time, thanks to the data balance method adopted, the proportion of mature and immature plums in the training set has changed from 2.2:1 to 1.2:1 so that the number of different categories of the dataset is similar.

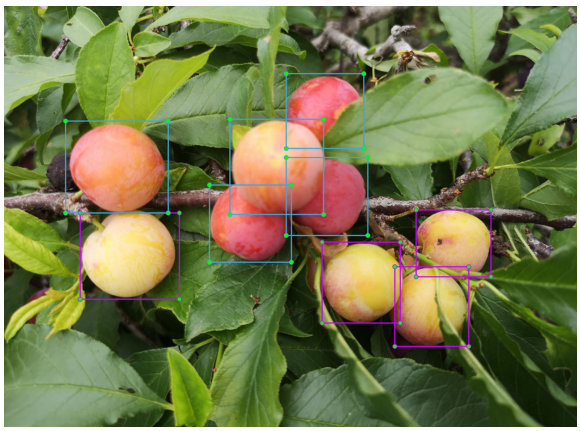


FIGURE 2 | Data annotation example: the blue box represents mature plums, and the purple box represents immature plums.

Methodologies

YOLOv4 Model

The YOLO series target detection models are widely used in industry and scientific research due to their excellent speed and detection accuracy performance. Bochkovskiy et al. (2020) proposed the YOLOv4 model based on YOLOv3, which has better recognition performance and faster speed. It can carry out end-to-end object prediction and classification. It is one of the most high-performance target detection methods at present. Compared with the YOLOv3 network, the main improvements of YOLOv4 include: (1) The Mosaic data augmentation method is designed, and the input images are merged by random clipping, scaling, and spatial arrangement. At the same time, training techniques, such as the learning rate cosine annealing attenuation method are used. (2) The new backbone network and activation function are used to enhance the feature learning ability of the network. Meanwhile, DropBlock regularization is used to alleviate the overfitting problem. (3) The Spatial Pyramid Pooling (SPP) module and Path Aggregation Network (PANet) structure are introduced. The PANet structure is used to transfer semantic features from top to bottom, and the feature pyramid is designed to transfer location features from bottom to top and aggregated through the backbone layer to improve the ability of network feature extraction. (4) The CIOU loss function is introduced to increase the width-to-height ratio information of the bounding box and enhance the robustness. The DIOU_nms prediction box screening mechanism is used to improve the screening performance of overlapping targets.

TABLE 1 | The number of datasets before and after augmentation.

Collection date	Dataset	Processing method	Number of pictures	Mature labels	Immature labels
April 24, 2021	Sub-dataset 1	Before augmentation	368	1,353	3,287
		After augmentation	4,416	16,236	39,444
May 2, 2021	Sub-dataset 2	Before augmentation	400	2,347	258
		After augmentation	2,400	9,388	1,548
May 3, 2021	Sub-dataset 3	Before augmentation	744	4,634	317
		After augmentation	4,464	27,804	1,902
	Total	Before augmentation	1,512	8,334	3,862
		After augmentation	11,280	53,428	42,894

A YOLOv4 network model mainly consists of the backbone, neck, and head networks. The backbone network is the CSPDarknet53 network, composed of 5 modules from Center and Scale Prediction 1 (CSP1) to CSP5, and each module is alternately stacked with CSPX and synthesis module of convolution, batch regularization, and Mish activation function (CBM) modules. After the input picture passes through the backbone network, the feature maps with three scales of $52 \times 52 \times 256$, $26 \times 26 \times 512$, and $13 \times 13 \times 1,024$ are obtained. The feature maps of different scales contain semantic information of different dimensions. For the $13 \times 13 \times 1,024$ feature layer, the maximum pooling of different scales is performed in the SPP structure to increase the receptive field of the network. After that, the three feature layers obtained are input into the PANet for a series of feature fusion, and finally, three detection heads of 13×13, 26×26, and 52×52 are output, respectively. Through decoding and non-maximum suppression of the detection head, the final prediction box is generated to detect the objects of different scales.

Depthwise Separable Convolution

Depthwise separable convolution is a lightweight convolution method, which can effectively reduce the amount of calculation compared with standard convolution. For the feature map with an input size of (D_x, D_y, M), the principle of depthwise separable convolution is to first separate Channel-by-channel convolve M convolution kernels of size (D_k, D_k) and each channel of the input feature map, and then, obtain a feature map where the input channel is equal to the output channel. Finally, N convolution kernels with size (1, 1) are used to pointwise convolution the feature map, and a new feature map (D_w, D_h, N) is obtained. Under the premise that the convolution characteristics are similar to the standard convolution performance, depthwise separable convolution can effectively reduce the network model's parameter amount and calculation amount. Furthermore, the speed of model training and reasoning is significantly accelerated.

Backbone Network

To pursue the model's high accuracy and better performance, many scholars have deepen the number of layers of the network model. However, this scheme has some drawbacks, such as increasing the number of parameters of the model, aggravating the calculation of the model, and reducing the operation efficiency of the model, which make it difficult to deploy on devices with limited computing resources. In picking robot operation, real-time performance is one of the most critical performance indicators, so it is necessary to lightweight the network reduce the calculation amount of the model. Although the CSPDarknet53 network used in the YOLOv4 model has strong feature extraction performance, the model is complicated and requires more computation.

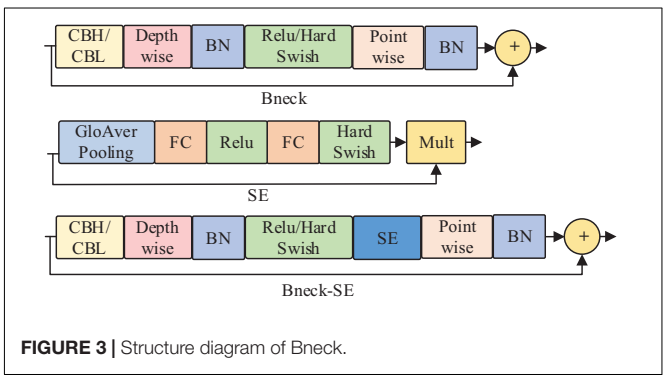


FIGURE 3 | Structure diagram of Bneck.

The MobileNetV3 network (Howard et al., 2019) combines deep separable convolution, MobileNetV2's inverted residual structure with linear bottleneck (Howard et al., 2018), and MnasNet's lightweight attention model based on the squeeze and excitation structure (Hu et al., 2018). MobileNetV3 constructs the network by combining these layers as construction Bneck, which successively passes through 1×1 ascending convolution, 3×3 depthwise separable convolution, and 1×1 dimension reduction convolution. The structure is shown in **Figure 3**. Moreover, the lightweight attention mechanism of the SE structure is introduced further to improve the feature extraction ability of the model. Eventually, the whole network structure is composed of Bneck stacks. Wherein CBL and CBH represent the synthesis modules of convolution, batch regularization, and LekeyReLU or h-swish activation functions; BN represents Batch Normalization; FC represents Full Connection; SE represents squeeze-and-excitation.

The Proposed Algorithm

To effectively identify dense plums, this paper chose 608×608 as the input size of the YOLOv4 model, and MobileNetV3 was used to replace the CSPDarknet53 backbone network of the original

model, which could effectively reduce the number of parameters of the model backbone network. The depthwise separable convolution was employed to replace the standard convolution in the original PANet to further reduce the number of model parameters. The model convolution module can obtain higher feature information through multiple down-sampling. However, when the feature layer with higher semantic information in the feature fusion network is up-sampled and fused, the convolution module will lose a certain amount of information, so the detection accuracy of small targets will be reduced. Therefore, this paper introduced the 152 × 152 × 24 layer to obtain more abundant shallow information to achieve fine-grained detection of small target objects. Due to the small pixels of plums in the whole image, the model will pay too much attention to the simple training samples and ignore the samples that are difficult to classify. Therefore, this paper introduced the Focal loss function to measure the contribution of difficult classification and easy classification to the total loss. The combined loss function of Focal loss and CIOU loss was designed as the loss function of the improved model. On this basis, this paper used transfer learning to train the model. Through the two-stage learning, the model's generalization performance can be further improved, and the dense plums can be identified quickly and accurately. The improved YOLOv4 model structure is shown in **Figure 4**. Among them, Conv means convolution, and DSC means depthwise separable convolution. DSC × 5 indicates that five depthwise separable convolution operations are required.

Multiscale Fusion Network Structure

In this study, the YOLOv4 algorithm was improved to solve the problem of insufficient feature extraction in dense plums recognition. When the input image size selected by the YOLOv4 model is 608 × 608, the feature layer responsible for predicting dense small targets is 76 × 76, and each feature grid's corresponding receptive field size is 8 × 8. When the input picture resolution is 1, 920 × 1, 080, the corresponding long edge is 25 through YOLO grid compression. That is to say; when the target feature size is less than 25 × 25 pixels, the target feature information cannot be effectively learned.

To extract the feature information of dense plums as much as possible, this study improved the network model of YOLOv4. Four feature layers were output from the backbone network MobileNetV3, namely P1 (152 × 152), P2 (76 × 76), P3 (38 × 38), and P4 (19 × 19). Among them, the P4 feature layer has the largest receptive field, which is suitable for large-scale target detection, and the receptive field of the P3 feature layer is suitable for medium-scale target detection. P2 is up-sampled and fused with the P1 feature layer, a relatively rich shallow layer can be obtained, which enables to achieve the fine-grained detection of small target objects. In the process of feature propagation, P4 is still obtained through the SPP structure. This study combines the feature layers P4, P3, P2, and P1 with different pyramid-level feature maps through up-sampling in the feature pyramid network (FPN) structure. Each feature layer is transformed by convolution and up-sampling to obtain the same scale and channel number as the previous feature layer and then stacked and fused with the previous feature layer to obtain a feature map with more abundant information. The improved network structure is shown in **Figure 4**.

The four feature layers from the FPN feature fusion output were pruned to prevent the network from being too redundant. The specific operation was that the 152 × 152 scale feature layer output by FPN is no longer the predicted output and directly up-sampled in the PANet structure. Therefore, the improved algorithm maintains the prediction YOLO head of three scales, namely P2' (76 × 76), P3' (38 × 38), and P4' (19 × 19).

Furthermore, the depthwise separable convolution was introduced into the PANet structure to replace the partial convolution of the original network. The improvement can effectively compress the number of network parameters and the amount of calculation.

Improvement of the Loss Function

The loss function of YOLOv4 consists of CIOU bounding box loss, classification loss, and confidence loss. The calculation method is shown in Formula (2)–(6):

$$L = L_{CIOU} + L_{class} + L_{conf} \tag{2}$$

$$L_{CIOU} = 1 - IOU(A, B) + \frac{\rho^2(A_{ctr}, B_{ctr})}{c^2} + \alpha\nu \tag{3}$$

$$\nu = \frac{4}{\pi^2}\left(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h}\right)^2 \tag{4}$$

$$\alpha = \frac{\nu}{(1 - IOU) + \nu} \tag{5}$$

$$IOU = \frac{|A \bigcap B|}{|A \bigcup B|} \tag{6}$$

Among them, A and B represent the area of the prediction frame and the actual frame, and the range of IOU is [0,1]; w^{gt} represents the width and height of the actual frame; w and h represent the width and height of the prediction frame; A_{ctr} and B_{ctr} represent the coordinates of the predicted box's center points and the actual box; ρ represents the Euclidean distance; c is the diagonal length of the smallest bounding box C composed of A and B; ν represents the penalty term.

Owing to the small physical size of plums and fewer pixels occupied in the image, when there are single, occluded, and densely stacked plums in an image, the model will automatically pay attention to and train single or easy-to-recognize simple samples, ignoring adhesion, and other difficult to classify samples. Therefore, it is necessary to find an appropriate loss function to balance the contribution of hard-to-classify and easy-to-classify samples to the total loss.

The Focal loss focused on hard-to-classify samples during the training process without affecting the original detection speed. Formula (7) of this function is as follows (Li et al., 2020; Long et al., 2021; Zhao et al., 2021):

$$\mathrm{FL}(p_t) = \begin{cases} -\alpha_t(1-p_t)^{\gamma}\ln(p_t), & \text{if } y = 1 \\ -(1-\alpha_t)\,p_t{}^{\gamma}\ln(1-p_t), & \text{otherwise} \end{cases} \tag{7}$$

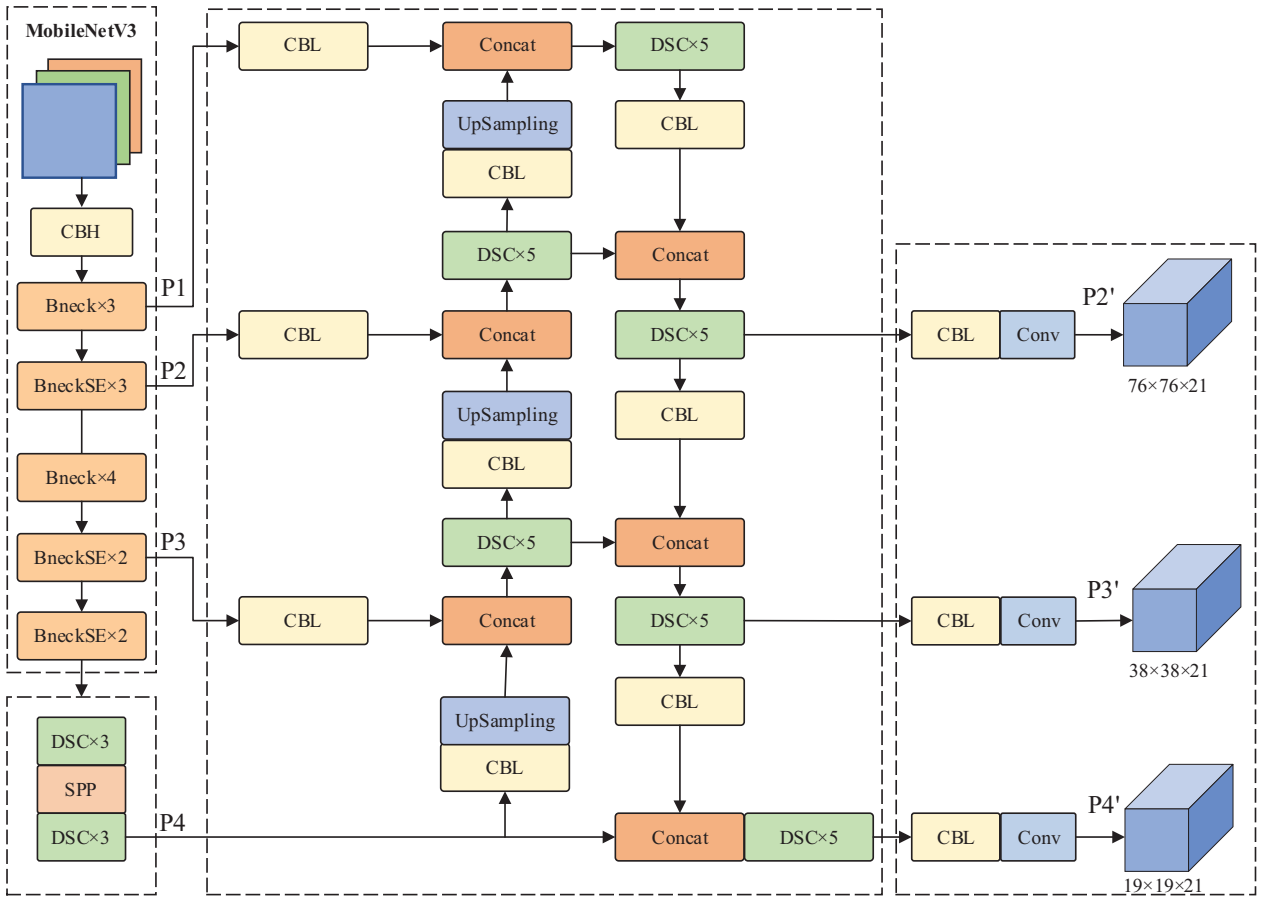


FIGURE 4 | The structure diagram of improved YOLOv4.

Where y is the number of sample labels; p_t represents the probability of belonging to the plum category; α_t is the coefficient of balancing the weight of positive and negative samples, $0 < \alpha_t < 1$; γ is the modulation parameter for complex samples.

This paper employed Focal Loss to replace class loss in the original loss function. Taking the prediction of simple mature plum as an example, when the p_t value is small, and the $(1-p_t)^{\gamma}$ value is close to 1, and its loss is almost unaffected. When p_t is large and close to 1, it indicates that the classification prediction result is better. If it is not corrected, it will easily interfere with the optimization direction of the model. After introducing Focal Loss, when p_t is larger, $(1-p_t)^{\gamma}$ is smaller. With the increase of γ, the faster the rate of simple sample reduction is adjusted, and the lower the proportion of simple samples in the total loss value. Therefore, the network model can focus more on hard-to-classify samples by introducing Focal Loss.

Plum Model Training Based on Transfer Learning

The hardware and software platform for model training was configured as follows: CPU is AMD R5-5600X 3.7 GHz, memory is 32 GB, storage SSD is 512 GB, display card is NVIDIA RTX2060S, display memory is 8 GB, the operating system is Windows10, CUDA version is 10.1, Python version is 3.7, and the PyTorch version is 1.6.

In this experiment, the input image pixels are $1,920 \times 1,440$. The K-means algorithm was used to generate the anchors' coordinate frame iteratively, and the Adam optimizer was used. The improved loss function was used to train the model. In addition to offline augmentation methods, Mosaic data augmentation was used in the training process to enrich the background of the detected objects further, strengthen the cognition of the network model on plum characteristics, and enhance the robustness and generalization performance of the model. The initial value of the learning rate was set to 10^{-4}, and the cosine annealing learning rate was optimized and updated during the training process.

To speed up the convergence of the model, this paper adopted the transfer learning method for training. The training was divided into two stages, and the whole stage was trained for 100 epochs. For the first half of the stage, the pre-training weight of the MobileNetV3 network was loaded, and the backbone feature extraction network of the model was trained 50 epochs by freezing. The initial value of the learning rate was set to 1×10^{-3}, and the batch size was set to 16. This operation can accelerate the convergence speed and prevent the pre-training

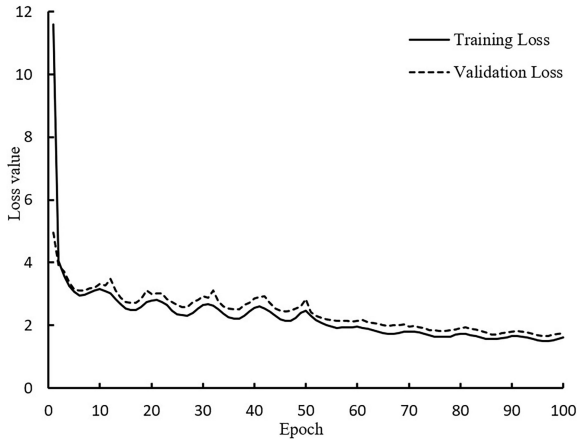


FIGURE 5 | Loss curve during training process.

weight from being destroyed. For the second half of the stage, the backbone feature extraction network was unfrozen, and the entire model was further trained for 50 epochs with an initial learning rate of 1×10^{-4}, and the batch size was set to 8. The convergence of the entire model was accelerated through two stages, and the training time of the model was shortened. In the training process, validation is performed after each epoch of training, and there is no overlap of the validation and test set. The weight file of each round of training was saved, and the loss values of the training set and validation set were saved. The loss value curves of the training set and validation set of the improved model in this paper are shown in **Figure 5**.

EXPERIMENTAL RESULTS AND COMPARATIVE ANALYSIS

Model Evaluation Indicators

To objectively measure the target detection effect of the model on dense plums, the precision (P), recall (R), harmonic average F1 value (F1), average precision (AP), mAP, the number of network parameters, the size of the weight, and the detection speed were used to evaluate the trained model. The Intersection over Union (IoU) value was 0.5 in the experiment. The calculation formulas of P, R, F1, AP, and mAP are shown in formulas (8–12).

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = \frac{2PR}{P + R} \tag{10}$$

$$AP = \int_0^1 P(R)dR \tag{11}$$

TABLE 2 | Comparison of recognition effect of the improved model before and after data balance.

Dataset types	Types Name	Plum AP	Raw_plum AP	mAP
Unbalanced data	A dataset	91.77%	80.23%	86.00%
Balanced data	B dataset	91.10%	86.34%	88.72%

$$mAP = \frac{\int_{q=1}^{Q} AP(q)}{Q} \tag{12}$$

Among them, *TP* represents the number of correctly detected plums; *FP* represents the number of misclassified plums; *FN* represents the number of missed plums; *F*1 represents the harmonic average of accuracy and recall. When *F*1 is closer to 1, the model is better optimized. *AP* represents the area composed of the *PR* curve and the coordinate axis. The higher the *AP* value is, the better the performance of the target detection algorithm is. The *mAP* represents the AP average of multiple categories, and its value represents the general detection performance of the algorithm for different categories.

Detection speed refers to the length of the model detection time, which was used to evaluate the real-time performance of the detection models. It is usually measured by the number of frames per second (FPS). The larger the FPS, the faster the model detection speed. FPS refers to the number of images processed per second in this paper.

Data Balance Comparison Experiments

This study selected the improved model based on YOLOv4 to train the plum data before and after the data balance. The same test set was selected to detect, and the evaluation index results are shown in **Table 2**. The data balance had little effect on the recognition rate of mature plums, which were both remained above 90%. Compared with the recognition rate of plums before data balance, the recognition rate of immature plums after balance increased by 6.11%, and the mAP of the test set also increased from 86 to 88.72%, with an increase of 2.72 percentage points. Overall, the recognition gap of plums with different maturity levels is alleviated, and the robustness of the model is enhanced.

Figure 6 shows the comparison of detection results before and after data balancing in different scenes, where A dataset represents the plum detection effect before data augmentation and B dataset represents the plum detection effect after data augmentation. By comparing the detection results before and after the data augmentation, we used the yellow frames to find out the missing plums in the (B, E, and H) image and marked the specific area in the original and the two types of detection images. Similarly, we used the blue frames to mark the specific areas where the plum was mistakenly detected.

A comprehensive comparison shows that the model after data balance has significantly improved the detection accuracy of immature plums, which indicates that the model's ability to identify small sample features has been strengthened by improving the ratio of mature and immature plums. Meanwhile, the data-balanced model has improved the misdetection

FIGURE 6 | The comparison of detection effect of plum images before and after improved data balance.

TABLE 3 | Comparison of detection results of different architectures.

Architecture	Plum AP	Raw_plum AP	mAP	Model size	Parameters	FPS
YOLOv4	88.99%	83.95%	86.47%	244 MB	61.38 M	20.03
YOLOv4-tiny	87.51%	81.71%	84.61%	22.4 MB	5.77 M	112
MobileNet-SSD	87.12%	79.23%	83.18%	24.7 MB	5.98 M	82.84
Improved YOLOv4	90.58%	86.54%	88.56%	54.05 MB	11.00 M	42.55

detection and missed detection of plums in scenes occluded by leaves and branches. In conclusion, the experimental results show the effectiveness of the data balance method.

TABLE 4 | Evaluation results of plum test set under different light conditions.

Light conditions	Classes	P	R	F1	mAP
Natural light	plum	90.32%	88.19%	0.89	94.53%
	raw_plum	89.41%	91.69%	0.91	
	mean value	89.87%	89.94%	0.9	
Side light	plum	88.29%	89.09%	0.89	94.86%
	raw_plum	93.07%	92.61%	0.93	
	mean value	90.68%	90.85%	0.91	
Back light	plum	90.14%	80.33%	0.85	86.75%
	raw_plum	92.36%	81.46%	0.87	
	mean value	91.25%	80.90%	0.86	

Comparative Experiments of Different Detection Methods

To evaluate the detection superiority of the improved model, the dataset made in this paper was trained by different target detection algorithms. After the training was completed, the test work was performed on the same testing sample sets. The AP, mAP value, model size, and detection speed of the four methods are shown in **Table 3**. Overall, the four models all had higher mAP for plums. Significantly, the improved YOLOv4 model was 1.59, 3.07, and 3.46 percentage points higher than the original YOLOv4, YOLOv4-tiny, and MobileNet-SSD, respectively, in the detection results of mature plums. Compared with the other three models, the improved YOLOv4 model increased by 2.59, 4.83, and 7.31 percentage points in the detection results of immature plums. Compared with the original YOLOv4 model, the improved YOLOv4 network model has a relatively simple structure, the model size of the improved YOLOv4 is compressed

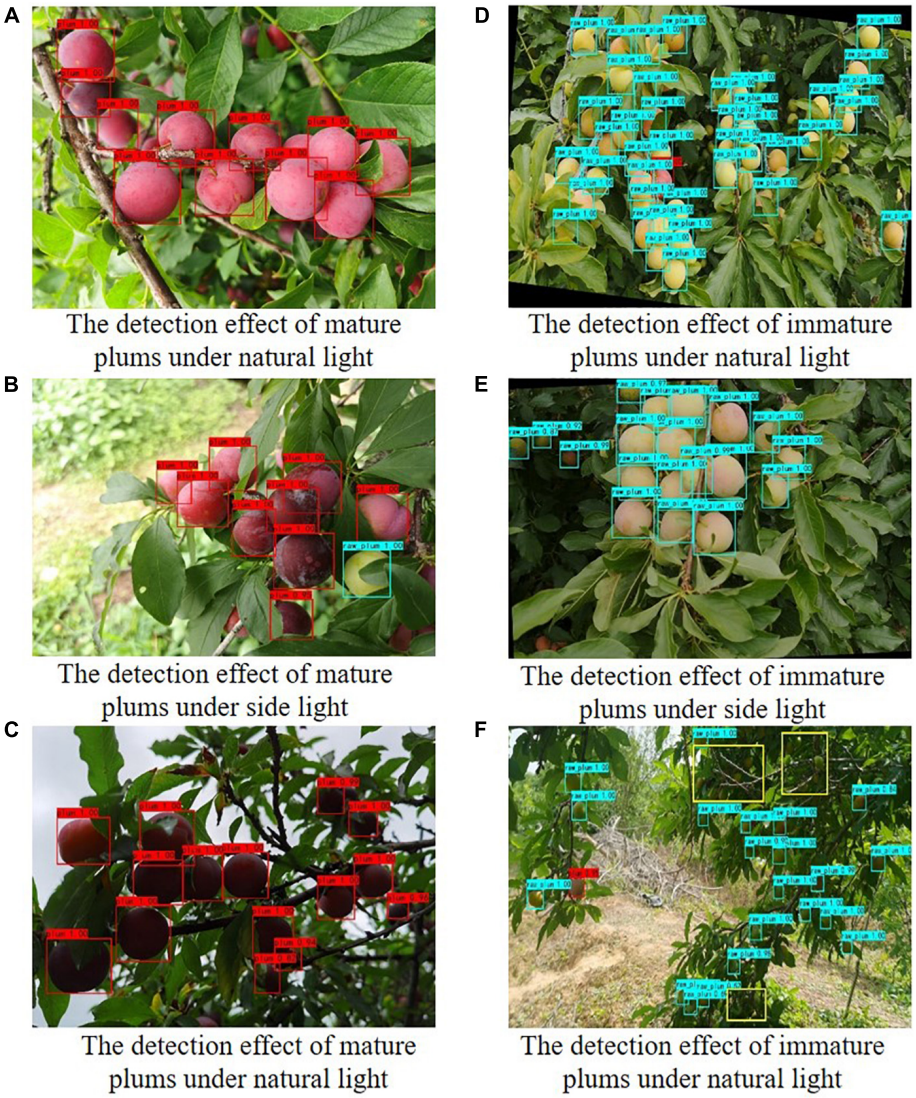


FIGURE 7 | Plum detection effect pictures under different light conditions.

by 77.85%, which is only slightly more than two times the combined model size of the YOLOv4-tiny and MobileNet-SSD. Moreover, the parameters is only 17.92% of the original YOLOv4's. The improved YOLOv4 network model is 112% faster than the original one in the terms of detection speed. In summary, the improved method presented in this paper shows the optimal detection performance for dense plums among the compared methods.

Comparative Experiment Under Different Light Conditions

The visual system of the fruit picking robot is susceptible to the influence of different lighting conditions in the natural environment when it collects videos or images, which affects the change of recognition accuracy. Under natural lighting conditions, the image is bright and dark, and plum contours are clear. Under backlight conditions, the overall image is dark, and plum contours are not evident. Under sidelight conditions, plums have uneven brightness. Therefore, 40 additional plum images were randomly selected under natural light, side light, and backlight to form a new test set C. The evaluation performance index results are shown in **Table 4**, and the detection results are shown in **Figure 7**.

TABLE 5 | The detection results of different density in four architectures.

Evaluation indicator	YOLOv4	YOLOv4-tiny	MobileNet-SSD	Improved YOLOv4
Moderately dense mAP value	89.19%	87.12%	87.28%	89.30%
Highly dense mAP value	83.01%	80.03%	77.16%	84.75%

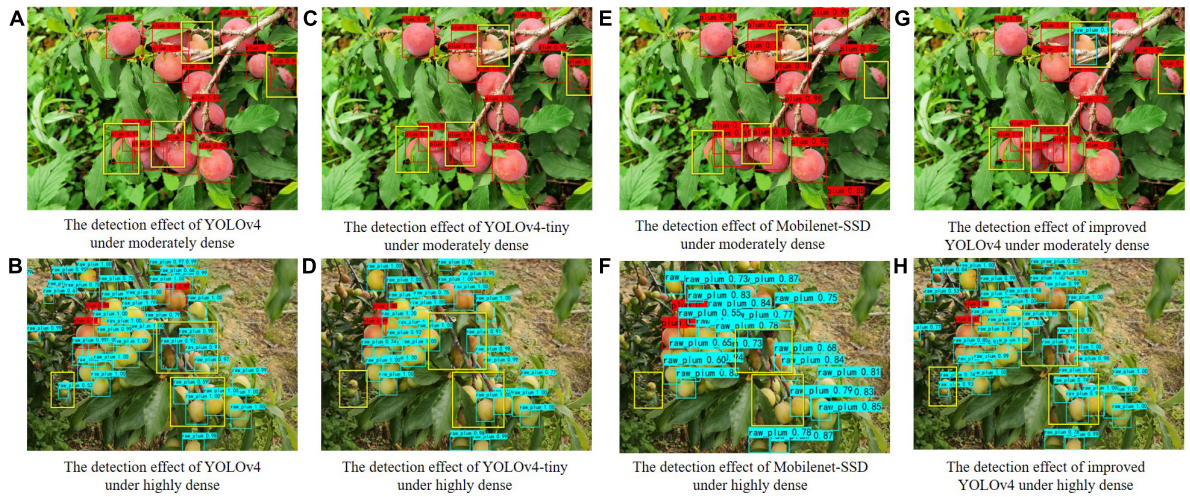


FIGURE 8 | Plum detection effect pictures under different dense conditions.

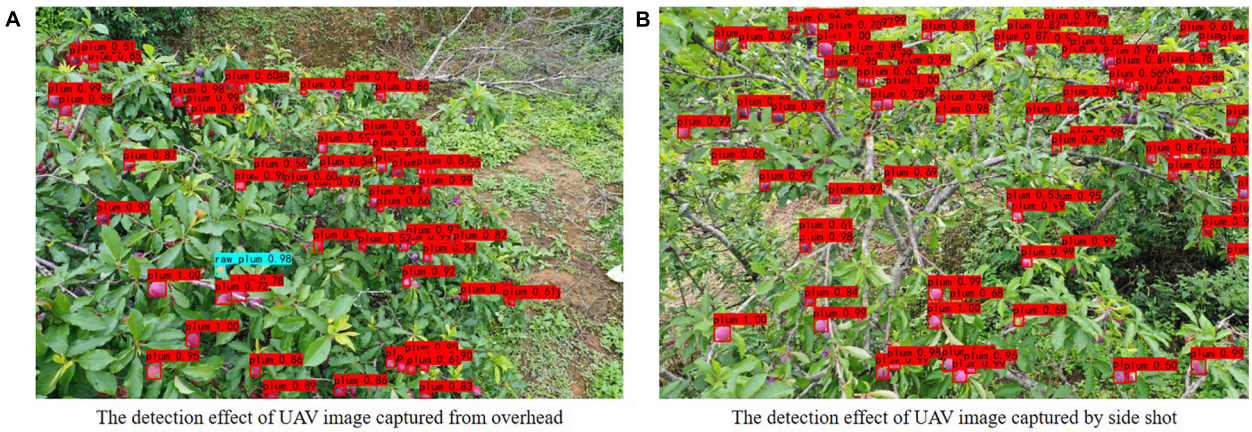


FIGURE 9 | The detection effect of unmanned aerial vehicle (UAV) images.

It can be seen from **Table 4** that the improved model maintains a high accuracy rate for plum detection results under different light, but the detection results are discrepant under different angle light conditions. Among them, the model has a slight decrease in performance under backlight conditions. Compared with natural light and side light conditions, the mAP value of the backlight is lower by 7.78 and 8.11%, respectively. Thanks to the clear texture of the plum under the conditions of natural and sidelight, the improved model can obtain higher recognition accuracy. However, the backlight condition has a certain interference effect on image feature extraction.

Figure 7 shows the comparison of the detection effects of plum images under different lighting conditions. It can be seen from **Figures 7A,D,B,E** that the plum has clear texture and uniform surface light intensity under natural light and sidelight. The difficulty of image detection is relatively small. Even the plum target at a distance can be detected. In the backlight, the image clarity is insufficient, and the color of mature plum fruit is dark red. Moreover, the color discrimination between immature plum and background (such as, branches or leaves) decreases, so a small amount of missing detection occurs. Overall, the improved model still maintains a high recognition accuracy in natural orchards.

Results and Analysis Under Dense Occlusion in Orchards

A Comparison Experiment of Plum Images With Different Density

We randomly selected some images with different densities for comparative experimental detection. If an image contains 10–20 plums, it is considered a moderately dense image. If there are more than 20 plums in the image, it is highly dense. Four architectures methods were used to test and compare the experimental results and detection results.

As can be known from **Table 5**, the accuracy of moderately dense plum images is higher than that of highly dense plum

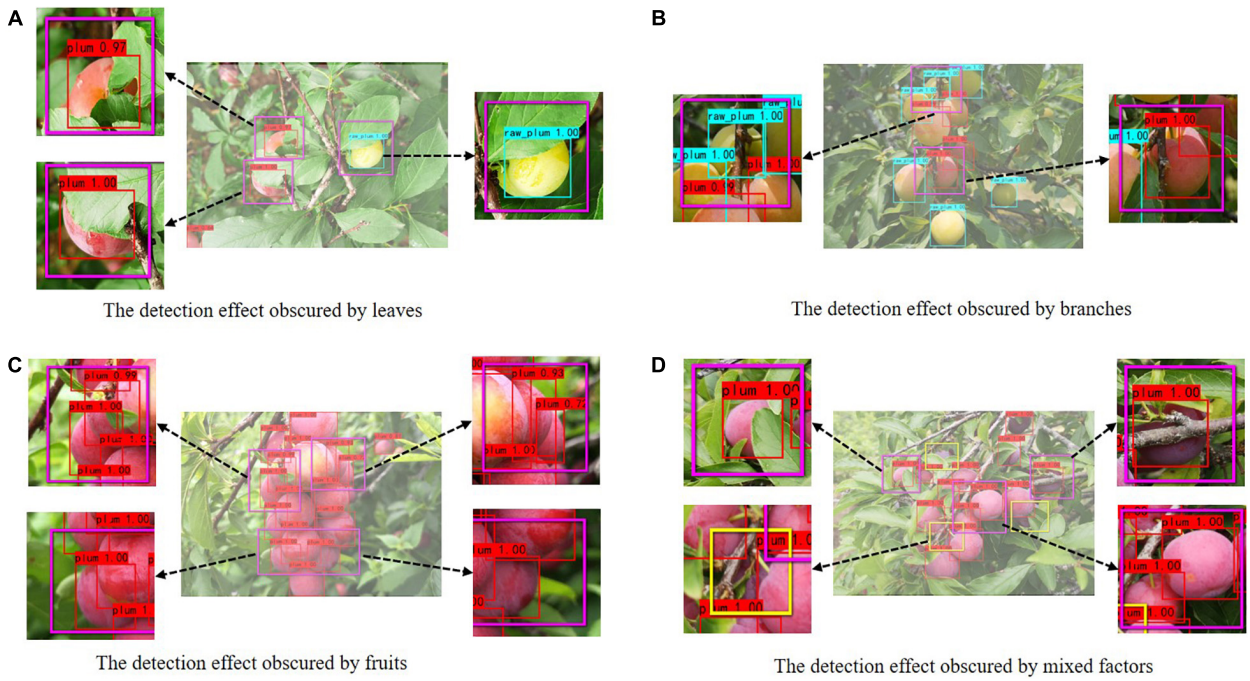


FIGURE 10 | The comparison of plum detection effect under different occlusions.

images, mainly due to the severe occlusion of highly dense plums, unclear fruit edges, and lack of texture features. By comparing the mAP of the four target detection models, the improved YOLOv4 has the highest mAP, the moderately dense recognition mAP reaches 89.30%, and the highly dense recognition mAP reaches 84.75%. The mAP gap between the two densities when compared showed that MobileNet-SSD has the largest mAP gap, exceeding 10%. The mAP gap of the improved model in the paper is the smallest, with a gap of only 4.55%. This shows that the improved model has a better detection effect for plums with different densities, and the improved model can narrow the detection gap of plums with different densities. Compared with other models, the improved model has a lower missed recognition rate and can recognize more plums, as shown in **Figure 8**. The experimental results show that the improved method in this paper has better detection accuracy, which indicates that the improved multiscale fusion structure can extract more valuable features under dense occlusion conditions.

To further explore the generalization ability of the improved model for image detection in a wide field of view, this study discussed the plum images from unmanned aerial vehicle (UAV) (DJI Yu2, zoom version) at a distance of 2–3 m from the tree canopy and 1–2 m parallel to the plum tree. Then, the improved model was employed to detect and evaluate the collected samples. The detection effect is shown in **Figure 9**. For the case of dense plums in a large field of view, plums can still be effectively identified by the improved model, indicating that the model has good generalization performance. The conclusion provides the possibility for further research on cooperative picking by UAV and ground fruit-picking robots.

A Comparative Experiment of Different Occlusion Situations

There may be some scenes obscured by branches, leaves, and other plums in the natural orchard. These occlusions may affect the detection accuracy of the model. For this reason, we also discussed the detection effect of the improved model on plum images with different occlusion categories.

The detection effect of the improved model for different occlusions is shown in **Figure 10**. The purple frame represents the partially enlarged image, and the yellow frame indicates the missed plums. As shown in **Figures 10A–C**, the model can efficiently recognize simple occlusion in the image. As shown in **Figure 10D**, when there is severe occlusion, plums with large area contour hidden or severely missing texture feature information will be missed. Nevertheless, on the whole, the improved model still has a good recognition effect, which indicates that the introduced Focal Loss function has a certain effect, making the model pay more attention to the occluded and difficult-to-recognize targets during the training process.

CONCLUSION

This study focused on dense plums in a real and complex orchard environment and proposed an improved YOLOv4 lightweight model. At first, the plums image data were collected, and the dataset was made using the automatic balancing method based

on category and the hybrid offline augmentation method. Then, MobileNetV3 and deep separable convolution were designed to improve the YOLOv4 model, and 152 × 152 feature layers were introduced to deal with the problem of insufficient feature extraction of the dense plums. Withal, the multiscale fusion and the joint loss function of Focal loss and CIOU loss were added to enhance the performance of the model against difficult-to-recognize plums. Finally, the improved model was trained by transfer learning. The main conclusions are as follows:

i. The accuracy of the data automatic balance algorithm proposed in this study for the detection of immature plum reached 86.34%, which is 6.11 percentage points higher than before the imbalance. The mAP increased from 86 to 88.72%, increasing 2.72 percentage points. Overall, the recognition gap of plums with different maturity levels is alleviated, and the robustness of the model is enhanced.
ii. Compared with the other three target detection models, the improved model based on YOLOv4 had the highest mAP result. By comparing with some results from the YOLOv4 model, the model size of the improved model is compressed by 77.85%, the total amount of parameters is only 17.92% of the original model parameters, and the detection speed is accelerated by 112%. The above data show that the improved model has achieved better performance in recognition accuracy and efficiency.
iii. This study discusses the detection performance of the improved model in natural scenes, such as different illuminations, different densities, images collected by UAV, and different occlusion conditions. The experimental results show that the improved model has excellent robustness and generalization performance.

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/s.

AUTHOR CONTRIBUTIONS

LW designed the experiments and wrote the manuscript. YZ and SL carried out the experiments. LW and YLi collected material data and analyzed experimental results with improved algorithms. SC and YLa supervised and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Laboratory of Lingnan Modern Agriculture Project (Grant No. NT2021009), Key-Area Research and Development Program of Guangdong Province (Grant No. 2019B020214003), the Leading Talents of Guangdong Province Program (2016LJ06G689), and Top Talents Program for One Case One Discussion of Shandong Province.

ACKNOWLEDGMENTS

We would like to thank the editor and reviewers of the Journal.

REFERENCES

Bochkovskiy, A., Wang, C., and Liao, H. (2020). Yolov4: optimal speed and accuracy of object detection. *arXiv* [Preprint] arXiv:2004.10934 [cs.CV],

Brown, J., and Sukkarieh, S. (2021). Dataset and performance comparison of deep learning architectures for plum detection and robotic harvesting. *arXiv* [Preprint] arXiv:2105.03832, [cs.CV],

Fu, L., Duan, J., Zou, X., Lin, G., Song, S., Ji, B., et al. (2019). Banana detection based on color and texture features in the natural environment. *Comput. Electron. Agric.* 167:105057. doi: 10.1016/j.compag.2019.105057

Fu, L., Duan, J., Zou, X., Lin, J., Zhao, L., Li, J., et al. (2020a). Fast and accurate detection of banana fruits in complex background orchards. *IEEE Access* 8, 196835–196846. doi: 10.1109/Access.2020.3029215

Fu, L., Majeed, Y., Zhang, X., Karkee, M., and Zhang, Q. (2020b). Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting. *Biosyst. Eng.* 197, 245–256. doi: 10.1016/j.biosystemseng.2020.07.007

Gao, J., Ni, J., Yang, H., and Han, Z. (2021). Pistachio visual detection based on data balance and deep learning. *Trans. CSAM* 52, 367–372. doi: 10.6041/j.issn.1000-1298.2021.07.040

Gao, X., Mo, M., Wang, H., and Leng, J. (2021). Recent advances in small object detection. *J. Data Acquis. Process.* 36, 391–417. (in Chinese with English abstract), doi: 10.16337/j.1004-9037.2021.03.001

Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, (Santiago), 1440–1448.

Gurubelli, Y., Ramanathan, M., and Ponnusamy, P. (2019). Fractional fuzzy 2DLDA approach for pomegranate fruit grade classification. *Comput. Electron. Agric.* 162, 95–105. doi: 10.1016/j.compag.2019.03.036

He, Z., Xiong, J., Chen, S., Li, Z., Chen, S., 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

Howard, A., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., et al. (2019). "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. (Seoul), 1314–1324. doi: 10.1109/ICCV.2019.00140

Howard, A., Zhmoginov, A., Chen, L., Sandler, M., and Zhu, M. (2018). "Inverted residuals and linear bottlenecks: mobile networks for classification, detection and segmentation," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (Salt Lake City, UT).

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2018). "Squeeze-and-Excitation Networks", in: IEEE trans. *Pattern Anal. Mach. Intell.* 2020, 7132–7141. doi: 10.1109/TPAMI.2019.2913372

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

Jang, E., Park, S., Park, W., Bae, Y., and Kim, H. (2021). 3D Image processing for recognition and size estimation of the fruit of plum (Japanese Apricot). *J. Korea Contents Assoc.* 21, 130–139. doi: 10.5392/JKCA.2021.21.02.130

Ji, W., Gao, X., Xu, B., Pan, Y., Zhang, Z., and Zhao, D. (2021). Apple target recognition method in complex environment based on improved YOLOv4. *J. Food Process Eng.* 44:e13866. doi: 10.1111/jfpe.13866

Kuznetsova, A., Maleva, T., and Soloviev, V. (2020). Using YOLOv3 algorithm with pre-and post-processing for apple detection in fruit-harvesting robot. *Agronomy* 10:1016. doi: 10.3390/agronomy10071016

Li, X., Qin, Y., Wang, F., Guo, F., and Yeow, J. (2020). "Pitaya detection in orchards using the mobilenet-YOLO model," in *Proceedings of the 2020 39th Chinese Control Conference (CCC)*, (Kunming), 6274–6278. doi: 10.23919/CCC50068.2020.9189186

Liao, W., Zheng, L., Li, M., Sun, H., and Yang, W. (2017). Green Apple Recognition in Natural Illumination Based on Random Forest Algorithm. *Trans. CSAM* 48, 86–91. (in Chinese with English abstract), doi: 10.6041/j.issn.1000-1298.2017.S0.014

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, 160–177. doi: 10.1007/s11119-019-09662-w

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, J., and Wang, X. (2020). Tomato diseases and pests detection based on improved yolo v3 convolutional neural network. *Front. Plant Sci.* 11:898. doi: 10.3389/fpls.2020.00898

Liu, T., Teng, G., Yuan, Y., Liu, B., and Liu, Z. (2021). Winter jujube fruit ecognition method based on improved YOLO v3 under natural scene. *Trans. CSAM* 52, 17–25. doi: 10.6041/j.issn.1000-1298.2021.05.002

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016). "SSD: single shot multibox detector", in: *Proceedings of the European Conference on Computer Vision-ECCV*. (Amsterdam). 816, 21–37. doi: 10.1007/978-3-319-46448-0_2

Long, Y., Li, N., Gao, Y., He, M., and Song, H. (2021). Apple fruit detection under natural condition using improved FCOS network. *Trans. CSAE* 37, 307–313. (in Chinese with English abstract), doi: 10.11975/j.issn.1002-6819.2021.12.035

Mo, J., Lan, Y., Yang, D., Wen, F., Qiu, H., Chen, X., et al. (2021). Deep learning-based instance segmentation method of litchi canopy from UAV-acquired images. *Remote Sens. Basel* 13:3919. doi: 10.3390/rs13193919

Pourdarbani, R., Sabzi, S., Hernandez-Hernandez, M., Hernandez-Hernandez, J., Garcia-Mateos, G., Kalantari, D., et al. (2019). Comparison of different classifiers and the majority voting rule for the detection of plum fruits in garden conditions. *Remote Sens. Basel* 11:2546. doi: 10.3390/rs11212546

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*, (Las Vegas, NV), 779–788. doi: 10.1109/CVPR.2016.91

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Las Vegas, NV), 7263–7271. doi: 10.1109/CVPR.2017.690

Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. *arXiv*[Preprint] 1–6. arXiv 1804.02767 [cs],

Ren, S., He, K., Girshick, R., and Sun, J. (2016). 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

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., and Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* 157, 417–426. doi: 10.1016/j.compag.2019.01.012

Wajid, A., Singh, N. K., Junjun, P., and Mughal, M. A. (2018). "Recognition of ripe, unripe and scaled condition of orange citrus based on decision tree classification," in *Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, (Sukkur), 1–4. doi: 10.1109/ICOMET.2018.8346354

Wang, X., and Liu, J. (2021a). Multiscale parallel algorithm for early detection of tomato gray mold in a complex natural environment. *Front. Plant Sci.* 12:620273. doi: 10.3389/fpls.2021.620273

Wang, X., and Liu, J. (2021b). Tomato anomalies detection in greenhouse scenarios based on YOLO-dense. *Front. Plant Sci.* 12:634103. doi: 10.3389/fpls.2021.634103

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.* 12:705021. doi: 10.3389/fpls.2021.705021

Wu, X., Qi, Z., Wang, L., Yang, J., and Xia, X. (2020). Apple detection method based on light-YOLOv3 convolutional neural network. *Trans. CSAM* 51, 17–25.

Xiong, J., Liu, Z., Tang, L., Liu, R., Bu, R., and Peng, H. (2018). Research on green citrus vision detection technology in natural environment. *Trans. CSAM* 49, 45–52. (in Chinese with English abstract), doi: 10.6041/j.issn.1000-1298.2018.04.005

Xue, Y., Huang, N., Tu, S., Mao, L., Yang, A., Zhu, M., et al. (2018). Immature mango detection based on improved YOLOv2. *Trans. CSAE* 34, 173–179. (in Chinese with English abstract), doi: 10.11975/j.issn.1002-6819.2018.07.022

Zhang, J., Karkee, M., Zhang, Q., Zhang, X., Yaqoob, M., Fu, L., et al. (2020). Multi-class object detection using faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting. *Comput. Electron. Agric.* 173:105384. doi: 10.1016/j.compag.2020.105384

Zhao, H., Qiao, Y., Wang, Y., and Yue, Y. (2021). Apple fruit recognition in complex orchard environment based on improved YOLOv3. *Trans. CSAE* 37, 127–135. (in Chinese with English abstract), doi: 10.11975/j.issn.1002-6819.2021.16.016

Zhao, Y., Gong, L., Zhou, B., Huang, Y., and Liu, C. (2016). Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis. *Biosyst. Eng.* 148, 127–137. doi: 10.1016/j.biosystemseng.2016.05.001

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpls.2022.850606

Tea Chrysanthemum Detection by Leveraging Generative Adversarial Networks and Edge Computing

Chao Qi[1†], Junfeng Gao[2†], Kunjie Chen[1], Lei Shu[1*] and Simon Pearson[2]*

[1] College of Engineering, Nanjing Agricultural University, Nanjing, China, [2] Lincoln Agri-Robotics Centre, Lincoln Institute for Agri-Food Technology, University of Lincoln, Lincoln, United Kingdom

A high resolution dataset is one of the prerequisites for tea chrysanthemum detection with deep learning algorithms. This is crucial for further developing a selective chrysanthemum harvesting robot. However, generating high resolution datasets of the tea chrysanthemum with complex unstructured environments is a challenge. In this context, we propose a novel tea chrysanthemum – generative adversarial network (TC-GAN) that attempts to deal with this challenge. First, we designed a non-linear mapping network for untangling the features of the underlying code. Then, a customized regularization method was used to provide fine-grained control over the image details. Finally, a gradient diversion design with multi-scale feature extraction capability was adopted to optimize the training process. The proposed TC-GAN was compared with 12 state-of-the-art generative adversarial networks, showing that an optimal average precision (AP) of 90.09% was achieved with the generated images (512 × 512) on the developed TC-YOLO object detection model under the NVIDIA Tesla P100 GPU environment. Moreover, the detection model was deployed into the embedded NVIDIA Jetson TX2 platform with 0.1 s inference time, and this edge computing device could be further developed into a perception system for selective chrysanthemum picking robots in the future.

Keywords: tea chrysanthemum, generative adversarial network, deep learning, edge computing, NVIDIA Jetson TX2

OPEN ACCESS

Edited by:
Daobilige Su,
China Agricultural University, China

Reviewed by:
Marcin Wozniak,
Silesian University of Technology, Poland
Saeed Hamood Alsamhi,
Ibb University, Yemen

***Correspondence:**
Kunjie Chen
kunjiechen@njau.edu.cn
Lei Shu
lei.shu@njau.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *07 January 2022*
Accepted: *09 March 2022*
Published: *07 April 2022*

Citation:
Qi C, Gao J, Chen K, Shu L and Pearson S (2022) Tea Chrysanthemum Detection by Leveraging Generative Adversarial Networks and Edge Computing. Front. Plant Sci. 13:850606. doi: 10.3389/fpls.2022.850606

INTRODUCTION

Some researches indicated that tea chrysanthemum has great commercial value (Liu et al., 2019). Besides, tea chrysanthemums offers a range of health benefits (Yue et al., 2018). For instance, it can considerably suppress carcinogenic activity and has significant anti-aging effects (Zheng et al., 2021). In the field, a tea chrysanthemum plant could present multiple flower heads, varying in different growth stages and sizes. Normally, tea chrysanthemums at the early flowering stage hold the best commercial value and health benefits, so they are mainly manually harvested at the early flowering stage, and this is a labor-intensive and time-consuming process.

Rapid developments in artificial intelligence and robotics offer a new opportunity to automate this harvesting task, dealing with the current scarcity of the skilled laborers (Dhaka et al., 2021; Kundu et al., 2021; Liu et al., 2021; Wieczorek et al., 2021). Hence, it is urgent to develop a selective harvesting robot. The perception system and manipulator are the two key components for developing selective harvesting robot. Many studies have shown that a high resolution image dataset has a profound impact on detection performance as it contains fine-grained features for object recognition (Zhou et al., 2021). However, collecting a dataset of tea chrysanthemums presents inherent difficulties. Tea chrysanthemums normally mature once a year and have to be

picked at the early flowering stage to maximize commercial values. Moreover, the early flowering stage is incredibly short, typically from only 2 days to 1 week. Currently, there is no publicly available dataset on tea chrysanthemums worldwide for developing a detection algorithm, which is a hindrance to build an intelligent selective harvesting robot and other intelligent phytoprotection equipment (Ansari et al., 2020; Alsamhi et al., 2021, 2022), e.g., Internet of Things based solar insecticidal lamps. Therefore, it is important to have a good dataset of tea chrysanthemums.

Using classical data augmentation to expand datasets and balance categories were reported in Tran et al. (2021). Nevertheless, classical data enhancement methods (rotation, translation, flipping, and scaling, etc.) only allow for restricted feature diversity, prompting the utilization of generated data. Generated samples provide more variation and further enrich the dataset to improve training accuracy. Recent approaches address the data generation issues through utilizing generative adversarial networks (GANs) (Wang et al., 2019). These methods use an encoder-decoder strategy to generate fake images that can be used to enrich the original dataset. GANs have shown the impressive results by generating stunning fake images such as human faces (Zhao et al., 2019). However, GANs still suffer from non-negligible flaws. In our case, three issues need to be further investigated.

Issue 1: In the current agricultural field, GAN generates images with a maximum resolution of 256 × 256 pixels. This is not suitable for the chrysanthemum detection task as the low resolution images contain restricted information about the environment related features, which somewhat affects the robustness of the whole model. How to generate images that can meet the detection task resolution of the tea chrysanthemum is an issue requiring further exploration.

Issue 2: The traditional GAN directly provides the latent code to the generative network, resulting in a massive feature entanglement, thus directly influencing the diversity of the generated chrysanthemum images. How to design a network structure that could improve the diversity of the generated chrysanthemum images is an issue to be further explored.

Issue 3: The alternating optimization of generators and discriminators makes the GAN prone to pattern collapse and gradient vanishing during training, so how to achieve stable training is an issue to be further explored.

Based on these three issues, we propose a tea chrysanthemum – generative adversarial network (TC-GAN) that can generate images with diversity at 512 × 512 resolution, as well as stable training. We decouple the latent code into intermediate vectors via a Mapping Network, resulting in controlling the diversity of chrysanthemum features. Also, we apply path length regularization in the Mapping Network, leading to more reliable and consistent behavior of the model and making architectural exploration easier. In the generative network, we add Stochastic variation after each convolutional layer to increase the diversity of the chrysanthemum images. Finally, we embed Res2Net into the generative network so that we can better guide the gradient flow to alleviate pattern collapse and gradient vanishing during the training process.

In this article, our goal is to generate datasets that can be used for the tea chrysanthemum detection task. We tested the images generated by TC-GAN on some state-of-the-art object detection models, as well as our own proposed detection model (TC-YOLO) (Qi et al., 2022). Moreover, for subsequent development work on an automated selective chrysanthemum picking robot, we chose to test the images generated by TC-GAN on a low-power embedded GPU platform, the NVIDIA Jetson TX2, as shown in **Figure 1**.

The contributions of this article are as follows:

1. High resolution (512 × 512) images of tea chrysanthemums with complex unstructured environments (illumination variations, occlusions, overlaps) were generated using the proposed TC-GAN model.
2. The images generated with TC-YOLO quantified the impact of five aspects, i.e., (1) dataset size, (2) epoch number, (3) different data enhancement methods, (4) various object detection models, and (5) complex unstructured environments on the TC-YOLO model, and verified the superiority of the TC-GAN model by comparing with some state-of-the-art GANs.
3. TC-YOLO, developed from images generated by TC-GAN, was successfully deployed and tested in the edge device NVIDIA Jetson TX2.

The rest of this article is organized as follows. Section "Related Work" describes the research background. Section "Materials and Methods" depicts the proposed TC-GAN structure. Section "Results" presents the experimental details. Section "Discussion" describes the contribution of this article and the limitations of the research, as well as pointing out possible future solutions. Section "Conclusion" gives a concise summary of this article.

RELATED WORK

Some GANs emerged to response the aforementioned Issue 3. Conditional Generative Adversarial Net (CGAN) (Liu et al., 2020) can strengthen the robustness of the model by applying conditional variables to the generator and discriminator that alleviate pattern collapse. Deep convolutional generative adversarial networks (DCGAN) (Jeon and Lee, 2021), the first GAN architecture based on convolutional neural networks, demonstrates a stable training process that effectively mitigates pattern collapse and gradient vanishing, but suffers from low quality and inadequate diversity of the generated images. Wasserstein GAN (WGAN) (Zhou et al., 2022) uses Wasserstein as an alternative to Jensen-Shannon (JS) divergence for comparing distributions, producing better gradients and improving training stability. Nevertheless, WGAN has difficulty converging due to the use of weight clipping, which can lead to sub-optimal performance.

Conditional Generative Adversarial Net, DCGAN, and WGAN had a profound impact on the development of GAN. Moreover, with the development of deep learning techniques, some high-performance GANs emerged to mitigate pattern

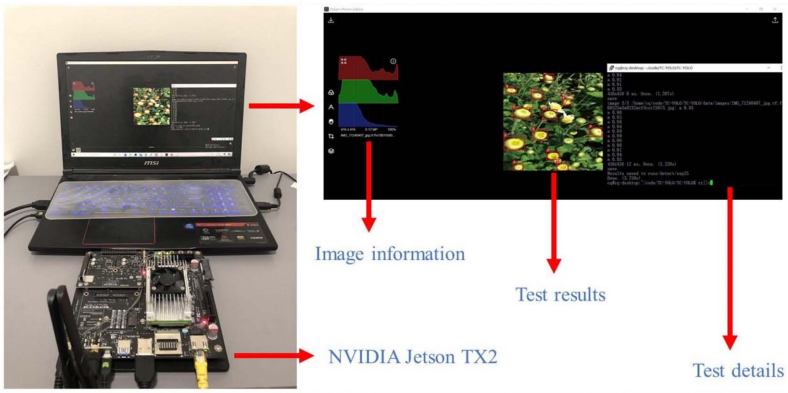


FIGURE 1 | The results of testing tea chrysanthemum – generative adversarial network (TC-GAN) on NVIDIA Jetson TX2. First, we used an HDMI cable to connect the laptop with the Jetson TX2, and ensure that the laptop and Jetson TX2 were under the same wireless network. Then, the TC-YOLO model and the tea chrysanthemum dataset were embedded in the flashed Jetson TX2 for testing.

TABLE 1 | Details of the twelve latest generative adversarial networks.

Algorithm	Published year	Characteristic	Resolution
Progressive GAN (Collier et al., 2018)	2017	Grow the generator and discriminator progressively	64 × 64
LSGAN (Mao et al., 2019)	2017	Applying the least squares loss function	112 × 112
SN-GAN (Mufti et al., 2019)	2018	Applying spectral normalization	32 × 32
MGAN (He et al., 2019)	2018	Applying multi-channel gait templates	64 × 64
Dist-GAN (Tran et al., 2018)	2018	Applying a latent-data distance constraint	64 × 64
Rob-GAN (Liu and Hsieh, 2019)	2019	Jointly optimize generator and discriminator	128 × 128
AutoGAN (Gong et al., 2019)	2019	Applying NAS algorithm	64 × 64
BigGAN (Qiao et al., 2020)	2018	Applying orthogonal regularization	512 × 512
Improved WGAN (Yang et al., 2020)	2020	Injecting an instance noise	128 × 128
Improved WGAN-GP (Kim et al., 2021)	2021	Wasserstein GAN with gradient penalty	28 × 28
Improved DCGAN (Chao et al., 2021)	2021	Applying batch normalization	64 × 64
DAG (Tran et al., 2021)	2021	Improve learning of the original distribution	48 × 48

collapse and gradient vanishing, resulting in stable training. Specific details are shown in **Table 1**. We will compare these models with the proposed TC-GAN in section "Results."

We collated the available literature on image recognition using GANs in agriculture, with a particular focus on the generated image resolution and the complex unstructured environment in the generated images, as shown in **Table 2**. High-resolution images contain better fine-grained features and more complex unstructured environments, facilitating the extraction of abundant image features for robust detection results. Also, high resolution images make transfer learning easier, and current object detection frameworks typically require datasets with a resolution higher than 416 × 416 (Liu and Wang, 2020). Not only that, to summarize the GANs in **Tables 1**, **2**, several structural improvements are needed. First, the latent codes (input vectors) in the GANs in **Tables 1**, **2** are directly fed into the generator network. Nevertheless, the design of using latent codes to generate specific visual features is somewhat restricted so that it has to consider the probability density of the input data. This design can prevent some latent codes from being mapped to features, resulting in feature entanglement. The proposed model structure allows vectors to be generated without considering the input data distribution through a custom mapping network, as well as reducing the correlation between different features. Second, multi-scale extraction and feature fusion can effectively guide the gradient flow, but in the GANs in **Tables 1**, **2**, the structure is designed mainly for normalization approaches, loss functions, control variables and mapping relationships between generators and discriminators. Currently, the structure of GANs lacks design for multi-scale extraction and feature fusion. The generator structure of the proposed model focuses on the combination of multi-scale extraction and feature fusion.

MATERIALS AND METHODS

Datasets

The tea chrysanthemum dataset utilized in this article was collected from October 2019 to October 2020 in Sheyang County, Dongzhi County and Nanjing Agricultural University, China. The datasets were all collected using an Apple X phone with an image resolution of 1080 × 1920. The datasets were captured in natural light under three environments, including illumination variation, overlap and occlusion. The chrysanthemums in the

TABLE 2 | Available literature using GAN for image recognition in agriculture.

Algorithm	Published year	Task	Accuracy (%)	Resolution	Test environment
DCGAN (Gandhi et al., 2018)	2018	Plant disease detection	88.6	64 × 64	Illumination
C-DCGAN (Hu et al., 2019)	2019	Tea leaf's disease identification	90	64 × 64	Illumination
DCGAN (Douarre et al., 2019)	2019	Apple scab segmentation	60	28 × 28	Ideal
CycleGAN (Padilla-Medina et al., 2019)	2019	Detection of apple lesions in orchards	95.57	64 × 64	Ideal
DCGAN (Bian et al., 2019)	2019	Tea clones identifications	76	64 × 64	Ideal
Deep CORAL (Marino et al., 2020)	2020	Potato defects classification	90	64 × 64	Ideal
CAAE (Zhong et al., 2020)	2020	Citrus plant diseases recognition	53.4	64 × 64	Illumination
DCGAN (Nafi and Hsu, 2020)	2020	Plant disease detection	86.63	64 × 64	Ideal
BEGAN (Luo et al., 2020)	2020	Pine cone detection	95.3	64 × 64	Ideal
CGAN (Olatunji et al., 2020)	2020	Kiwi geometry reconstruction	75	28 × 28	Ideal
DCGAN (Talukdar, 2020)	2020	Plant disease classification	95.88	64 × 64	Ideal
DCGAN (Hu et al., 2020)	2020	Recognition of diseased pinus trees	78.6	64 × 64	Ideal
TasselGAN (Shete et al., 2020)	2020	Plant traits detection	94	128 × 128	Illumination
CycleGAN (Zhao et al., 2021a)	2021	Bale detection	93	64 × 64	Ideal
DCGAN (Espejo-Garcia et al., 2021)	2021	Weeds identification	93.23	64 × 64	Ideal
DoubleGAN (Zhao et al., 2021b)	2021	Plant disease detection	99.06	64 × 64	Ideal
AR-GAN (Nazki et al., 2020)	2020	Plant disease recognition	86.1	256 × 256	Illumination

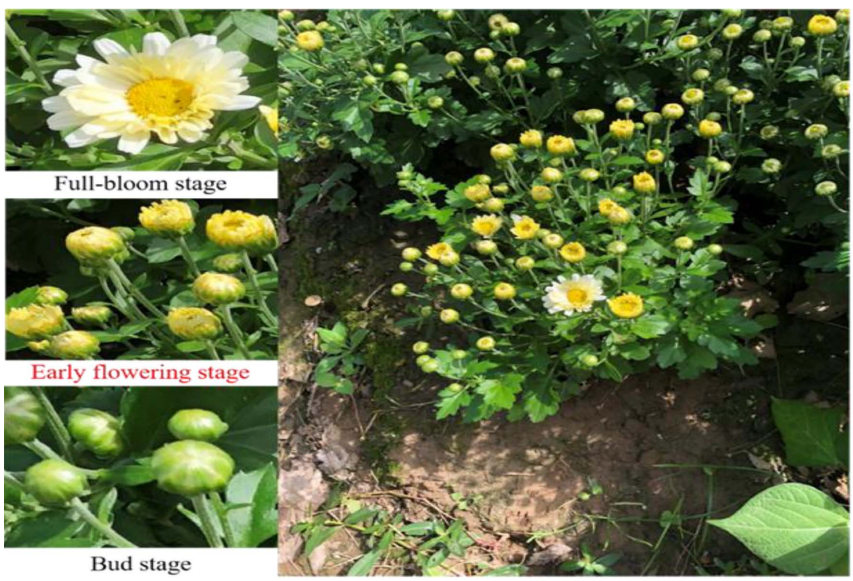


FIGURE 2 | Examples of the collected original images.

dataset comprise three flowering stages: the bud stage, the early flowering stage and the full bloom stage. The bud stage refers to when the petals are not yet open. The early flowering stage means when the petals are not fully open and the full bloom stage denotes when the petals are fully open. The three examples of the original images are shown in **Figure 2**.

NVIDIA Jetson TX2

There is no need to transmit all gathered image data back to cloud for further processing since the communication environment in countryside is generally not stable and the long time delay for smart equipment, i.e., chrysanthemum picking robot, is not acceptable. The NVIDIA Jetson TX2 has a 6-core ARMv8 64-bit CPU complex and a 256-core NVIDIA Pascal architecture GPU. The CPU complex consists of a dual-core Denver2 processor and a quad-core ARM Cortex-A57, as well as 8 GB LPDDR4 memory and a 128-bit interface, making it ideal for low power and high computational performance applications. Thus, this edge computing device was chosen to design and implement a real-time object detection system. We introduced the NVIDIA Jetson TX2 in **Figure 3**.

Architecture

The proposed TC-GAN comprises a generator and a discriminator. In the generator, the non-linear mapping network

JETSON TX2	Parameters
GPU	256 Core Pascal 1.3 TFLOPS (FP16)
CPU	6 core Denver and A57 (2x) 2MB L2
Memory	Up to 8 GB 128b LPDDR4 58 GB/s
Storage	Up to 32 GB eMMC
Encode	4K @ 60 (H.265)
Decode	2x 4K @ 60 (H.265)
Camera	12 lanes MIPI CSI-2 D-PHY 1.2 (30 Gbps) C-PHY (41 Gbps)
Mechanical	87mm x 50mm 400 pin connector

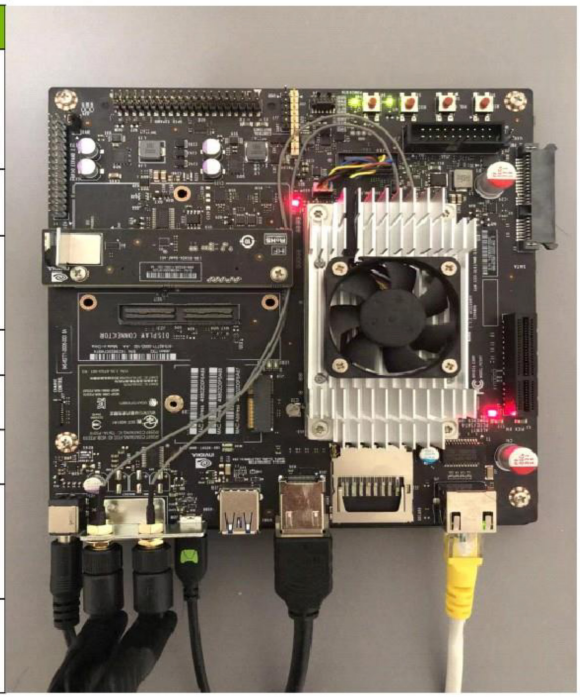


FIGURE 3 | NVIDIA Jetson TX2 parameters.

f is implemented with a 4-layer multilayer perceptron (MLP), as well as applying path length regularization to decorrelate neighboring features for more fine-grained control of the generated images. The learned affine transform then specializes w to the style y = (ys, yb), controlling the Adaptive Instance Normalization (AdaIN) operation after each convolutional layer of the synthetic network g, followed by Res2Net to better guide the gradient flow without increasing the network computational workload. Finally, we introduce noisy inputs that enable the generator to provide random detail. We inject a specialized noise image into each layer (4^2–512^2) of the generator network, these are single channel images composed of Gaussian noise. The noise images are used with a feature scaling factor broadcast to all feature maps, and subsequently applied to the output of the corresponding convolution. Leaky ReLU is employed as the activation function throughout the generator. In the discriminator, the generated 512 × 512 resolution image and the real image of the same resolution are fed into the discriminator network simultaneously and mapped to 4 × 4 via convolution. In the whole convolution process, some diverse modules are inserted, including CL (Convolution + Leaky ReLU) and CBL (Convolution + Batch Normalization + Leaky ReLU). It is worth noting that the GAN training tends to be unstable, and no extra modules are inserted to guide the gradient flow and make the overall discriminator network look as simple as possible. Also, due to the lack of gradient flow in the underlying layer, the BN module was not inserted in the convolution process. Leaky ReLU is utilized as the activation function throughout the discriminator. Moreover, the generator and discriminator both employ the Wasserstein distance with gradient penalty as the loss function. The structure of TC-GAN is shown in **Figure 4**.

Mapping Network

The mapping network consists of four fully connected layers that map the latent space z to the intermediate latent space w via affine transformations. **Figure 4** depicts the structure of the mapping network. To capture the location of latent codes with rich features, this network encourages feature-based localization. A mixed regularization strategy is adopted, where two random latent codes are used instead of one latent code to generate some images during the training process. When generating an image, we simply switch from one latent code to another at a randomly picked point in the generative network. Specifically, the two latent codes $z1$, $z2$ are under control in the mapping network, and the corresponding $w1$, $w2$ are allowed to fix the features so that $w1$ works before the intersection point and $w2$ works after the intersection point. This regularization strategy prevents neighboring features from being correlated. Furthermore, extracting potential vectors in a truncated or reduced sample space helps to improve the quality of the generated images, although a certain degree of diversity in the generated images would be lost. Based on this, we can consider a similar approach. First, after training, intermediate vectors are generated in the mapping network by randomly selecting different inputs and calculating the center of mass in these vectors:

$$\bar{w} = E_{z \sim P(z)} \tag{1}$$

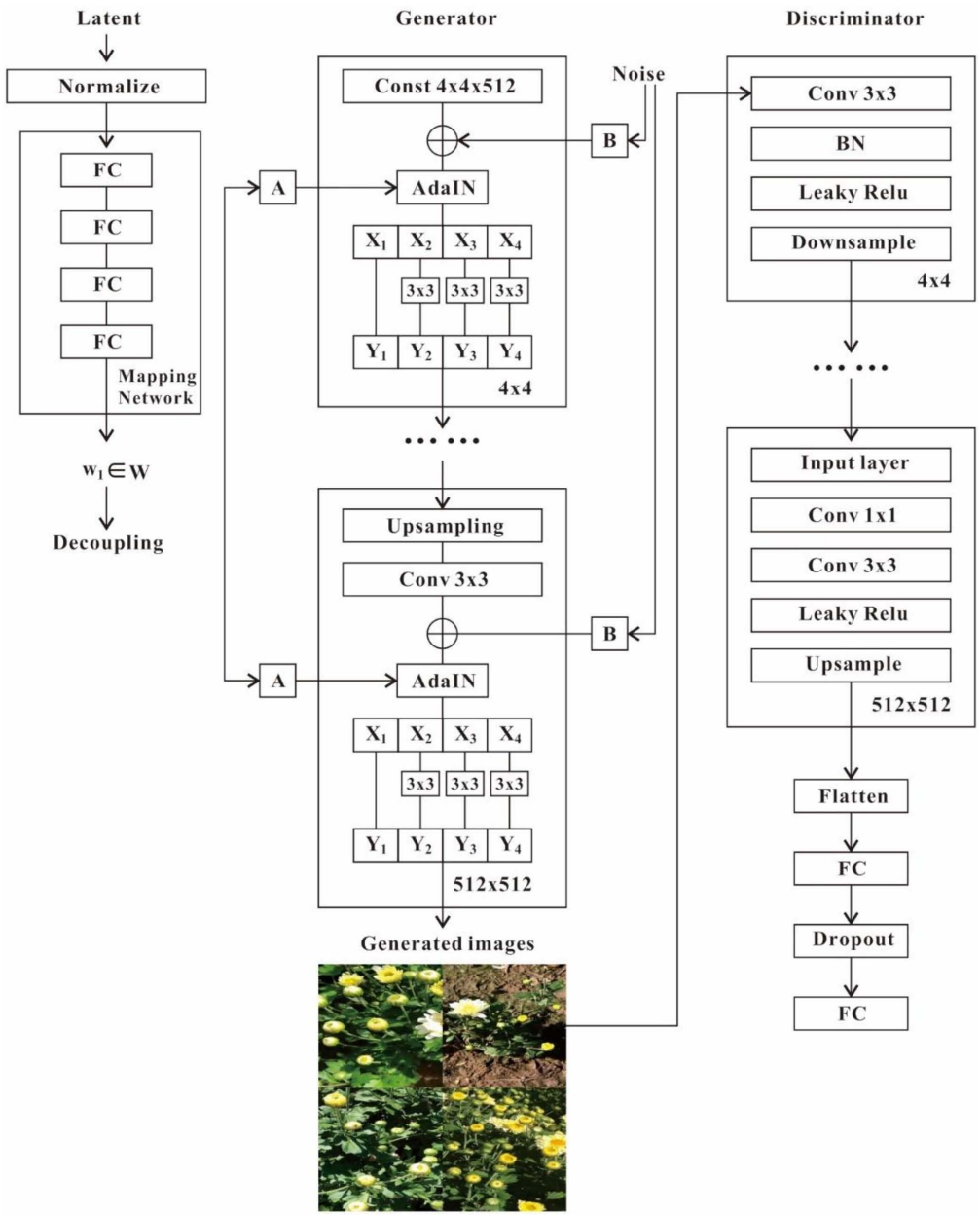


FIGURE 4 | Structure of the proposed TC-GAN network. Mapping network can effectively capture the location of potential codes with rich features, benefiting the generator network to accurately extract complex unstructured features. A represents the learned affine transformation. B denotes the learned per-channel scaling factor applied to the noisy input. Discriminator network is designed to guide the training of the generator network, which is continuously confronted by alternating training between the two networks, ultimately enabling the generator network to better execute the generation task.

where $\bar{w}$ stands for the center of mass and z denotes the latent space.

We can then scale the deviation of a given w from the center as:

$$w' = \bar{w} + \psi(w - \bar{w}) \quad (2)$$

where w' refers to the truncated w and ψ defines the difference coefficient between the intermediate vector and the center of mass.

Stochastic Variation

The sole input of traditional networks is through the input layer, which generates spatially varying pseudo-random numbers from earlier activations. This method consumes the capacity of the network and thus makes it difficult to hide the periodicity of the generated signal, causing the whole generation process unstable. To address this challenge, we embed noise along each convolutional layer. In a feature-based generator network, the entire feature map is scaled and biased with the same

values. As a result, global effects like shape, illumination or background style could be controlled consistently. Moreover, noise is applied to each pixel individually and thus is eminently suitable for controlling random variations. Once the generative network attempts to control the noise, this leads to spatially inconsistent decisions that will be penalized by the discriminator. Accordingly, TC-GAN can learn to use global and local channels properly without clear guidance.

Path Length Regularization

Path length regularization makes the network more reliable and makes architectural exploration easier. Specifically, we stimulate fixed-size steps of W to generate non-zero fixed-size variations in the image. The bias is measured by observing the corresponding gradient of W in the random direction, which should have a similar length regardless of w or the image space direction. This indicates that the mapping from potential space to image space is conditional.

At a single $w \in W$, the local metric scaling properties of the generator mapping $g(w)$: $W \rightarrow Y$ are fixed by the Jacobian matrix $J_w = \partial g(\mathrm{w}) / \partial \mathrm{w}$. Since we wish to preserve the expected length of the vector regardless of its direction, we formulate the regularizer as:

$$E_{w,y\sim N(0,I)}(\| J_w^T y \|_2 - a)^2 \tag{3}$$

where y is a random image with normally distributed pixel intensities, and $w \sim f(z)$, where z are normally distributed. In higher dimensions, this prior is minimized when J_w is orthogonal at any w. An orthogonal matrix retains length and does not introduce squeezing across any dimension.

This prior is minimized when the expected value of y reaches the minimum at each latent space point w, respectively, and we start from the internal expectation:

$$\mathcal{L}_w := E_y \left(\| J_w^T y \|_2 - a \right)^2 \tag{4}$$

We use the single-valued decomposition $J_w^T = U \tilde{\Sigma} V^T$ for analysis. Where $U \in R^{L \times L}$ and $V \in \mathbb{R}^{M \times M}$ represent orthogonal matrices. Since rotating a unit normal random variable by an orthogonal matrix will make its distribution invariant, the equation simplifies to:

$$\mathcal{L}_w = E_y \left(\| U \tilde{\Sigma} V^T y \|_2 - a \right)^2 = E_y \left(\| \tilde{E} y \|_2 - a \right)^2 \tag{5}$$

Moreover, the zero matrix effectively marginalizes its distribution in dimension. Then, we simply consider the minimization of the expression:

$$\mathcal{L}_w = E_{\tilde{y}} \left(\| \Sigma \tilde{y} \|_2 - a \right)^2 \tag{6}$$

where $\tilde{y}$ is a unit-normal distribution in dimension L. All matrices J_W^T that share the same singular values as Σ generate the same raw loss values. When each diagonal entry of the diagonal matrix Σ is given the specific same value, thus writing the expectation into the integral of the probability density over $\tilde{y}$:

$$\mathcal{L}_{\mathrm{w}} = \int \left(\| \Sigma \tilde{\mathrm{y}} \|_2 - a \right)^2 p_{\tilde{\mathrm{y}}}(\tilde{\mathrm{y}}) \mathrm{d}\tilde{\mathrm{y}}$$

$$= (2\pi)^{-\frac{L}{2}} \int \left(\| \Sigma \tilde{\mathrm{y}} \|_2 - a \right)^2 \exp\left(-\frac{\tilde{\mathrm{y}}^T \tilde{\mathrm{y}}}{2} \right) \mathrm{d}\tilde{\mathrm{y}} \tag{7}$$

To observe the radially symmetric form of the density, we alter to polar coordinates $\tilde{y} = r\phi$. Such a variable change is replaced by the Jacobian factor r^{L-1}:

$$\tilde{\mathcal{L}}_w = (2\pi)^{-\frac{L}{2}} \int_{\mathbb{S}} \int_0^{\infty} (r \| \Sigma \phi \|_2 - a)^2 r^{L-1} exp\left(-\frac{r^2}{2} \right) dr d\phi \tag{8}$$

where r represents the distance from the origin, and ϕ stands for a unit vector. Thus, the $(2\pi)^{-L/2} r^{L-1} \exp\left(-\frac{r^2}{2}\right)$ denotes the L-dimensional unit average density expressed in polar coordinates. The Taylor approximation argument indicates that when L is high, the density is well-approximated by density $(2\pi e/L)^{-\frac{L}{2}} \exp\left(-\frac{1}{2}(r-\mu)^2/\sigma^2\right)$for any ϕ. Replacing the density into the integral, the loss is given by approximately:

$$\mathcal{L}_{\mathrm{w}} \approx (2\pi e/L)^{-L/2} \int_{\mathrm{S}} \int_0^{\infty} (r \| \Sigma \phi \|_2 - a)^2 \exp\left(-\frac{(r-\sqrt{L})^2}{2\sigma^2} \right) dr \mathrm{d}\phi \tag{9}$$

where the approximation turns out to be exact in the limit of infinite dimension L.

By minimizing this loss, we set Σ to obtain a minimum of the function $(r \| \Sigma \phi \|_2 - a)^2$ over a spherical shell of radius $\sqrt{L}$. According to this function becoming constant in ϕ, the equation reducing to:

$$\mathcal{L}_w \approx (2\pi e/L)^{-L/2} \mathcal{A}(S) a^2 L^{-1} \int_0^{\infty} (r - \sqrt{L})^2 exp\left(-\frac{(r-\sqrt{L})^2}{2\sigma^2} \right) dr \tag{10}$$

where $\mathcal{A}(S)$ indicates the surface area of the unit sphere.

To summarize, we proved that, supposing a high dimensionality L of the latent space, the path length prior at each latent space point w is minimal if all the singular values of the Jacobian matrix for the generator are equal to a global constant, that is, they are orthogonal up to a global constant. We avoid the explicit computation of the Jacobian matrix by using the same $J_w^T y = \nabla w \left(g(w) \cdot y \right)$, and this could be efficiently computed by standard back-propagation. The constant a is dynamically set to a long-term exponential moving average of length $\| J_w^T y \|_2$, enabling the optimization to discover the appropriate global scale on its own.

Res2Net

To alleviate pattern collapse and gradient vanishing, we use a gradient diversion approach (Res2Net) with stronger multi-scale feature extraction capabilities. In essence, a set of 3×3 filters are substituted with smaller filter groups, connected in a similar way to the residual mechanism. **Figure 5** illustrates Res2Net, we split the feature map uniformly into s subsets of feature maps after 1×1 convolution, denoted by x_i, where $i \in \{1, 2, \ldots s\}$. Each

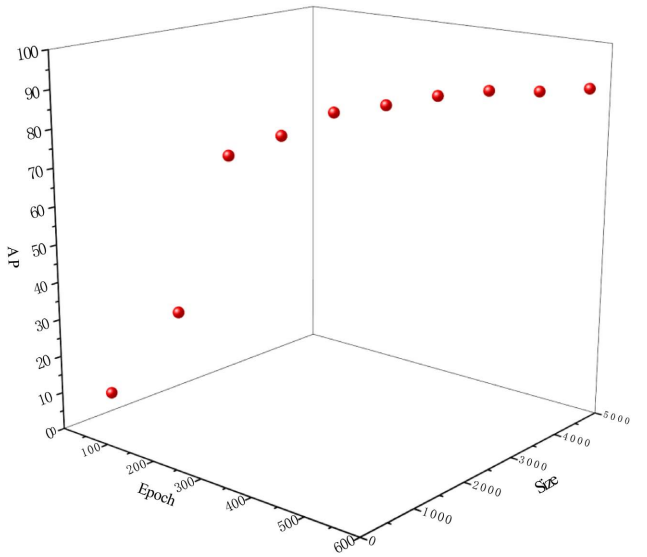


FIGURE 5 | Impact of dataset size and epoch time on TC-GAN.

subset of features x_i has the same spatial size compared to the input feature map, but with 1/s number of channels. Besides x_1, each x_i has a corresponding 3 × 3 convolution, denoted by K_i (). We denote the output of K_i () by y_i. This feature subset is summed with the output of K_{i-1} () and fed into K_i (). To minimize the parameters and increase s simultaneously, we skip the 3 × 3 convolution of x_1. Hence, y_i could be written as:

$$y_i = \begin{cases} x_i & i = 1; \\ K_i(x_i) & i = 2; \\ K_i(x_i + y_{i-1}) & 2 < i \leq s \end{cases} \quad (11)$$

Each 3 × 3 convolutional operator K_i () has the potential to capture feature information from feature splits $\{x_j,\ j \leq i\}$. When the feature slice x_j is passed through the 3 × 3 convolution operator, the output may have a larger receptive field than x_j . Due to the combinatorial explosion effect, the output of the Res2Net module contains varying amounts and various combinations of receptive field sizes.

In Res2Net, the global and local information of the chrysanthemum image is extracted through processing the splits in a multi-scale approach. To better fuse feature information at different scales, we tandem all the splits and compute them by 1 × 1 convolution. The segmentation and tandem approach allow for efficient convolution operations and feature processing. To minimize the parameter capacity, we skip the convolution of the first segmentation. In this article, we employ s to control parameters for the scale dimension. Larger s has the potential to allow learning features with richer perceptual field dimensions, with negligible computation of tandem.

Evaluation Metrics

Average precision (AP) is a common evaluation metric in object detection tasks. In this article, we calculate the average precision (IoU = 0.5) of the tea chrysanthemum to test the performance of the model. The equation is as follows:

$$AP = \sum_{k=1}^{N} P(k)\ \Delta recall\ (k) \quad (12)$$

where N represents the size of the test dataset, $P(k)$ stands for the precision value of the k tea chrysanthemum images, and recall (k) denotes the change in recall between k and k-1 tea chrysanthemum images.

In addition, error and miss rates were introduced in section "Impact of Different Unstructured Environments on the TC-YOLO" to investigate the ability of TC-GAN to generate unstructured environments. error rate indicates a ratio of the number of falsely detected samples to the total samples. miss rate refers to the ratio of undetected samples to the total samples.

Experimental Setup

The experiments were conducted on a server with an NVIDIA Tesla P100, CUDA 11.2. We built the proposed model using python with the pytorch framework. During training, the key hyperparameters were set as follows: epoch = 500; learning rate = 0.001; and the optimizer used was Adam.

RESULTS

Performance of Tea Chrysanthemum – Generative Adversarial Network in Datasets of Different Sizes

To verify the effect of the generated dataset size and the number of training epochs on the chrysanthemum detection task, we randomly selected the datasets with 10 different number of training samples (100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, and 4500) and corresponding ten different training epochs at 100, 200, 250, 300, 350, 400, 450, 500, 550, and 600, respectively, from the generated chrysanthemum dataset and tested them on the proposed TC-YOLO, the results are shown in **Figure 5**.

It can be seen that the performance of TC-YOLO improves with the increase of the dataset size and training epochs. When the dataset size is less than 1500 and the training epoch is less than 300, the AP value increases rapidly as the dataset size and the training epochs increase (13.54–80.53%, improved by 494.76%). When the dataset size reached 2500, and the training epoch reached 400, the AP values only slightly improved and finally converged (from 87.29 to 90.09%) with the increase of the number of samples and the training epochs. After the dataset size reached 4000 and the training epoch reached 550, the detection performance AP value decreased slightly to 89.51%. Combining these results, we set the optimal dataset size to 3500 and the optimal training epoch to 500 for the test experiments in Sections *B*, *C*, and *D*, as it achieved the highest AP values with the smallest dataset size and the least training epoch.

Study on the Performance of Traditional Data Enhancement Methods and Tea Chrysanthemum – Generative Adversarial Network

To investigate the performance of classical data enhancement methods and TC-GAN, we selected nine classical data enhancement methods and TC-GAN (**Table 3**). These data enhancement methods were configured and tested in the TC-YOLO object detection model. The results are shown in **Table 3**. TC-GAN shows the best performance with an AP value of 90.09%. It was surprising that the advanced data enhancement methods, such as Mixup, Cutout and Mosaic, had a disappointing performance with AP values of only 80.33, 81.86, and 84.31%, respectively. This may be due to the fact that a large amount of redundant gradient flow would greatly reduce the learning capacity of the network. We also found that the performance of Flip and Rotation was second only to TC-GAN, with AP values of 86.33 and 86.96%. The performance of the model improves slightly, with an AP value of 87.39% when Flip and Rotation are both configured on TC-YOLO. Even so, its AP is still 2.7% lower than TC-GAN.

Comparisons With State-of-the-Art Detection Models

To verify the superiority of the proposed model, tea chrysanthemum dataset generated by TC-GAN was used to compare TC-YOLO with nine state-of-the-art object detection frameworks (Kim et al., 2018; Zhang et al., 2018; Cao et al., 2020; Zhang and Li, 2020), and the results are shown in **Table 4**.

Table 4 shows that TC-GAN not only achieves excellent performance on the TC-YOLO object detection model with a mAP of 90.09%, but also performs well on other state-of-the-art object detection frameworks. TC-GAN is a general data enhancement method and not constrained to the specific object detectors. Generally speaking, large image sizes benefit model training by providing more local feature information, however, large image sizes ($>512 \times 512$) do not always result in improved performance. In **Table 4**, all the models with large image sizes ($>512 \times 512$) were unable to achieve a performance above 87%. The main reason for this may be that the image size generated in this article is 512×512, which would affect the performance of models requiring a large input size. To match the input size, the images could only be artificially resized to the smaller images,

TABLE 4 | Comparisons with state-of-the-art detection methods.

Method	Backbone	Size	FPS	mAP
RetinaNet	ResNet101	800 × 800	4.54	82.62
RetinaNet	ResNet50	800 × 800	5.31	80.59
RetinaNet	ResNet101	500 × 500	7.23	79.13
RetinaNet	ResNet50	500 × 500	7.87	83.68
EfficientDetD6	EfficientB6	1280 × 1280	5.29	81.23
EfficientDetD5	EfficientB5	1280 × 1280	6.21	83.51
EfficientDetD4	EfficientB4	1024 × 1024	7.93	83.19
EfficientDetD3	EfficientB3	896 × 896	9.28	84.83
EfficientDetD2	EfficientB2	768 × 768	11.66	84.22
EfficientDetD1	EfficientB1	640 × 640	15.26	82.93
EfficientDetD0	EfficientB0	512 × 512	37.61	82.81
M2Det	VGG16	800 × 800	7.08	80.63
M2Det	ResNet101	320 × 320	16.89	85.16
M2Det	VGG16	512 × 512	21.22	80.88
M2Det	VGG16	300 × 300	42.53	78.24
YOLOv3	DarkNet53	608 × 608	12.14	86.52
YOLOv3 (SPP)	DarkNet53	608 × 608	15.66	83.89
YOLOv3	DarkNet53	416 × 416	43.25	84.13
PFPNet (R)	VGG16	512 × 512	24.35	82.41
RFBNetE	VGG16	512 × 512	21.54	77.37
RFBNet	VGG16	512 × 512	45.46	85.53
RefineDet	VGG16	512 × 512	31.33	81.12
RefineDet	VGG16	448 × 448	43.31	79.66
YOLOv4	CSPDarknet53	608 × 608	19.22	85.11
YOLOv4	CSPDarknet53	512 × 512	24.63	84.34
YOLOv5l	CSPDenseNet	416 × 416	42.24	88.83
YOLOv5m	CSPDenseNet	416 × 416	36.91	86.68
YOLOv5x	CSPDenseNet	416 × 416	32.28	84.02
YOLOv5s	CSPDenseNet	416 × 416	47.88	88.29
TC-YOLO	CSPDenseNet	416 × 416	47.53	90.09

TABLE 3 | Performance comparison of different data enhancement methods.

Flip	Shear	Crop	Rotation	Grayscale	Blur	Mixup	Cutout	Mosaic	TC-GAN	AP
√										86.33
	√									84.21
		√								83.99
			√							86.96
				√						82.09
					√					80.13
						√				80.33
							√			81.86
								√		84.31
√			√							87.39
									√	90.09

√ *means that this data enhancement method has been adopted.*

resulting in a reduction in image resolution, and this would considerably affect the final test performance of the models. Also, transfer learning ability varies between models, and this may account for some models with over 512 × 512 resolutions performing poorly. Given the above two reasons, TC-YOLO has relatively better transfer learning ability compared to other object detection models. Therefore, TC-YOLO is used as the test model for generating chrysanthemum images in this article. Besides, TC-YOLO requires the image input size of 416 × 416, making the image resolution a relatively minor impact on the final performance. Furthermore, we deployed the trained TC-YOLO in the NVIDIA Jetson TX2 embedded platform to evaluate its performance for robotics and solar insecticidal lamps systems development. **Figure 6** shows the detection results.

Impact of Different Unstructured Environments on the TC-YOLO

Datasets with complex unstructured environments can effectively improve the robustness of detection models. This study investigated the ability of the proposed TC-GAN to generate complex unstructured environments, including strong light, weak light, normal light, high overlap, moderate overlap, normal overlap, high occlusion, moderate occlusion and normal occlusion, as shown in **Figure 7**. A total of 26,432 chrysanthemums were at the early flowering stage in the nine unstructured environments. Since there are no mature standards to define these different environments, we set the criteria based on empirical inspection. Strong light is defined as when sunlight obscures more than fifty percent of the petal area. Weak light is defined as when the shadows cover less than fifty percent of the pedal area. Normal light is defined as when the sunlight covers between zero and fifty percent of the petal area. High overlap is defined as when the overlapping area between petals is greater than sixty percent. Moderate overlap is defined when the overlapping area between petals is between thirty to sixty percent. Normal overlap is defined when the overlapping area between petals is between zero to thirty percent. High occlusion is defined as more than sixty percent of the petal area is obscured. Moderate is defined as when thirty to sixty percent of the petal area is obscured. Normal occlusion is defined as when zero to thirty percent of the petal area is obscured. The chrysanthemums are counted separately in different environments. For example, when chrysanthemums in normal light, normal overlap and normal occlusion appear in one image simultaneously, their numbers increase by one in the calculation.

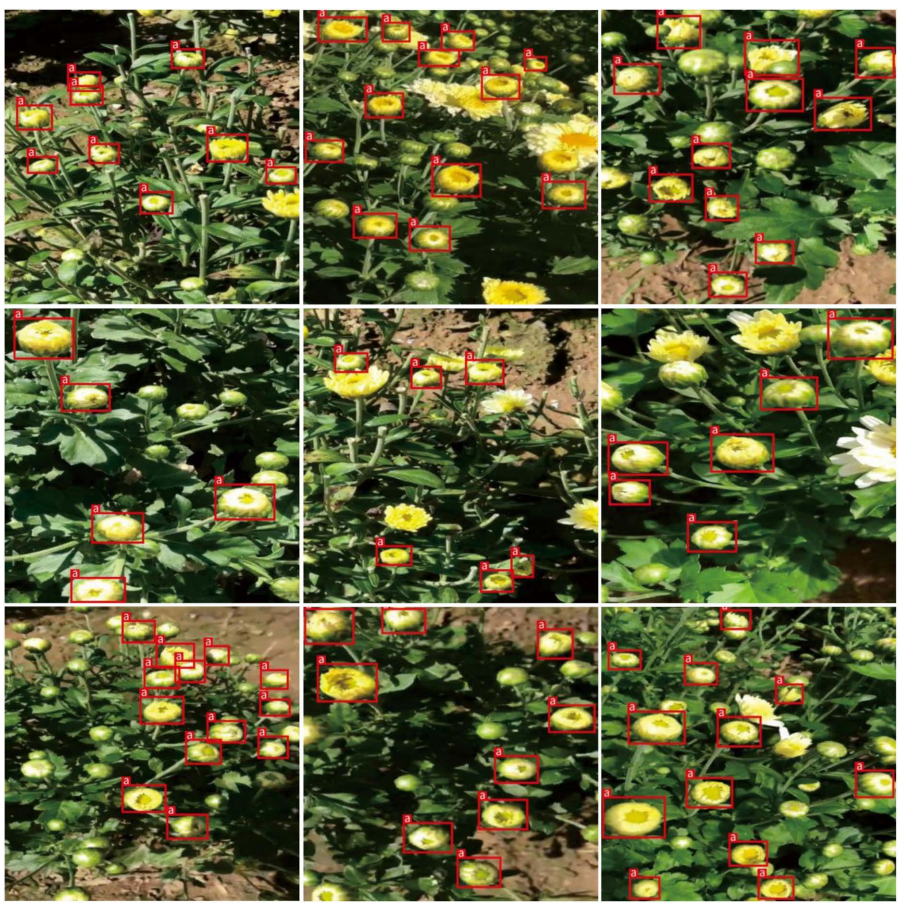


FIGURE 6 | Qualitative results of our method. The red box indicates the recognised tea chrysanthemum.

FIGURE 7 | Example of nine unstructured scenarios.

TABLE 5 | Impact of different unstructured scenarios on the TC-YOLO.

Environment	Count	Correctly identified		Falsely identified		Missed	
		Amount	Rate (%)	Amount	Rate (%)	Amount	Rate (%)
Strong light	6511	5021	77.12	686	10.54	804	7.25
Weak light	10162	8786	86.46	857	8.43	519	5.11
Normal light	18686	17458	93.43	988	5.29	240	1.28
High overlap	5249	4167	79.39	379	7.22	703	13.39
Moderate overlap	11892	10420	87.62	659	5.54	813	6.84
Normal overlap	17443	16499	94.59	419	2.4	525	3.01
High occlusion	7811	6284	80.45	729	9.33	798	10.22
Moderate occlusion	12162	10661	87.66	630	5.18	890	7.16
Normal occlusion	19299	18147	94.03	648	3.36	504	2.61

Table 5 shows that under normal conditions, with normal light, normal overlap and normal shading, the AP values reached at 93.43, 94.59, and 94.03%, respectively. When the unstructured environment became complicated, the AP values dropped significantly, especially under the strong light environment, with only 77.12%. AP value. Intriguingly, the error rate (10.54%) was highest under the strong light, probably because the light added shadows to the chrysanthemums. It also may be due to the poor ability of TC-GAN to generate high quality images under light environment. The high overlap had the highest miss rate of 13.39%. Furthermore, overall, overlap had the least influence on the detection of chrysanthemums at the early flowering stage. Under high overlap, the AP, error and miss rates were 79.39, 7.22, and 13.39%, respectively. Illumination had the biggest effect on chrysanthemum detection at the early flowering stage. Under high light, the accuracy, error and miss rates were 77.12, 10.54, and 7.25%, respectively.

Comparison of the Latest Generative Adversarial Neural Networks

To fully investigate the performance of TC-GAN, TC-GAN and 12 state-of-the-art generative adversarial neural networks were tested on the chrysanthemum dataset using the TC-YOLO model. The proposed TC-GAN generated chrysanthemum images with a resolution of 512 × 512. However, there is variability in the resolution of the generated images from different generative adversarial neural networks. Therefore, to facilitate testing of the TC-YOLO model and to ensure a fair competition between TC-GAN and these generative adversarial neural networks, we modified the output resolution of the latest generative adversarial neural networks. According to the original output resolution of these neural networks, we modified the output resolution of LSGAN, Improved WGAN-GP to 448 × 448, BigGAN kept the original resolution unchanged, and the output resolution of the remaining generative adversarial neural networks were all

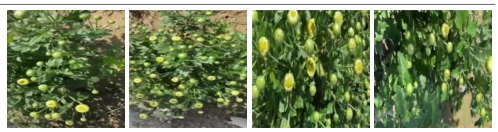
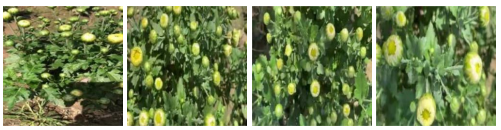
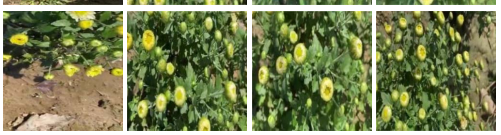
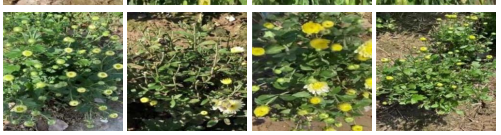
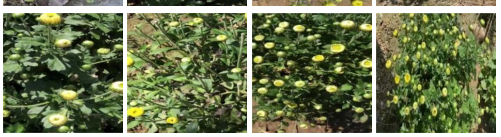
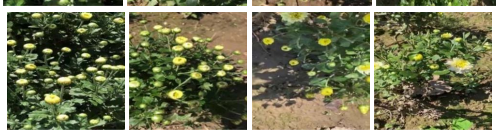
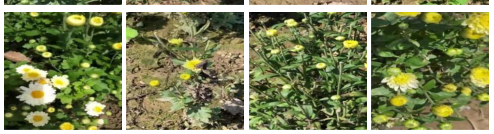
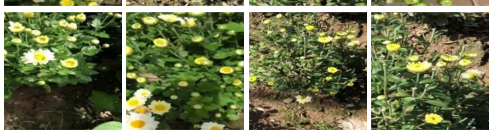
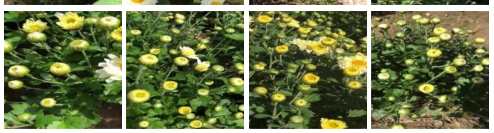
TABLE 6 | Comparison between tea chrysanthemum – generative adversarial network (TC-GAN) and state-of-the-art GANs.

Method	Size	Times/min	AP
Improved SN-GAN	32 × 32	1290	80.61
BigGAN	512 × 512	1610	86.45
Dist-GAN	64 × 64	1322	80.68
Progressive GAN	64 × 64	1256	81.11
LSGAN	112 × 112	1410	84.03
Rob-GAN	128 × 128	1293	85.28
MGAN	64 × 64	1151	82.39
AutoGAN	64 × 64	1340	83.25
Improved DCGAN	64 × 64	1280	84.38
DAG	48 × 48	1768	83.29
Improved WGAN-GP	28 × 28	1640	76.16
Improved WGAN	128 × 128	1501	87.16
TC-GAN	512 × 512	1460	90.09

adjusted to 512 × 512, while other parameters were kept fixed. The performance is shown in **Table 6**.

Table 6 shows some experimental details. TC-GAN has the best performance among the latest 12 generative adversarial neural networks, with an AP value of 90.09%. It is worth noting that TC-GAN does not have an advantage in training time among all the latest generative adversarial neural networks, with all nine models training faster than TC-GAN. Only BigGAN, Improved WGAN-GP and Improved WGAN are slower than TC-GAN, with training times of 50, 180, and 241 min slower than TC-GAN, respectively. This may be due to the design of the network structure, which increases the depth of the network and adds a gradient penalty mechanism. In contrast to most convolutional neural networks, deepening the structure of generative adversarial neural networks tends to make training unstable. Also, the gradient penalty mechanism is very sensitive to the choice of parameters, and this helps training initially, but subsequently becomes difficult to optimize. Furthermore, in general, the smaller the original generated image size, the worse the performance of the generative adversarial neural network in the detection task. This is because, firstly, current mainstream adversarial neural networks generate images with low resolution, and artificially enlarging the resolution would blur the image, thus affecting the detection accuracy. Then, some latest models, such as Progressive GAN, Improved DCGAN and so on, are designed for better faces, and these models are not robust in terms of transfer ability. Interestingly, among the 12 latest generative adversarial neural networks, most of the network structures are unconditional. Nevertheless, from a comprehensive performance perspective, network structures with conditional mechanisms, such as the improved WGAN, have surprisingly good performance. Its training time is only 41 min slower than TC-GAN, while the AP value is only slightly lower by 2.93%. Network structures with conditional mechanisms are undoubtedly valuable to learn from, and adding conditional mechanisms could be a future direction to improve the performance of TC-GAN. To visualize the performance of TC-GAN, the images generated by TC-GAN are shown in **Table 7**.

TABLE 7 | Generation results of different GANs.

Methods	Result
Improved WGAN-GP	
SN-GAN	
Dist-GAN	
Progressive GAN	
MGAN	
AutoGAN	
DAG	
LSGAN	
Improved DCGAN	
Rob-GAN	
BigGAN	
Improved WGAN	
TC-GAN	

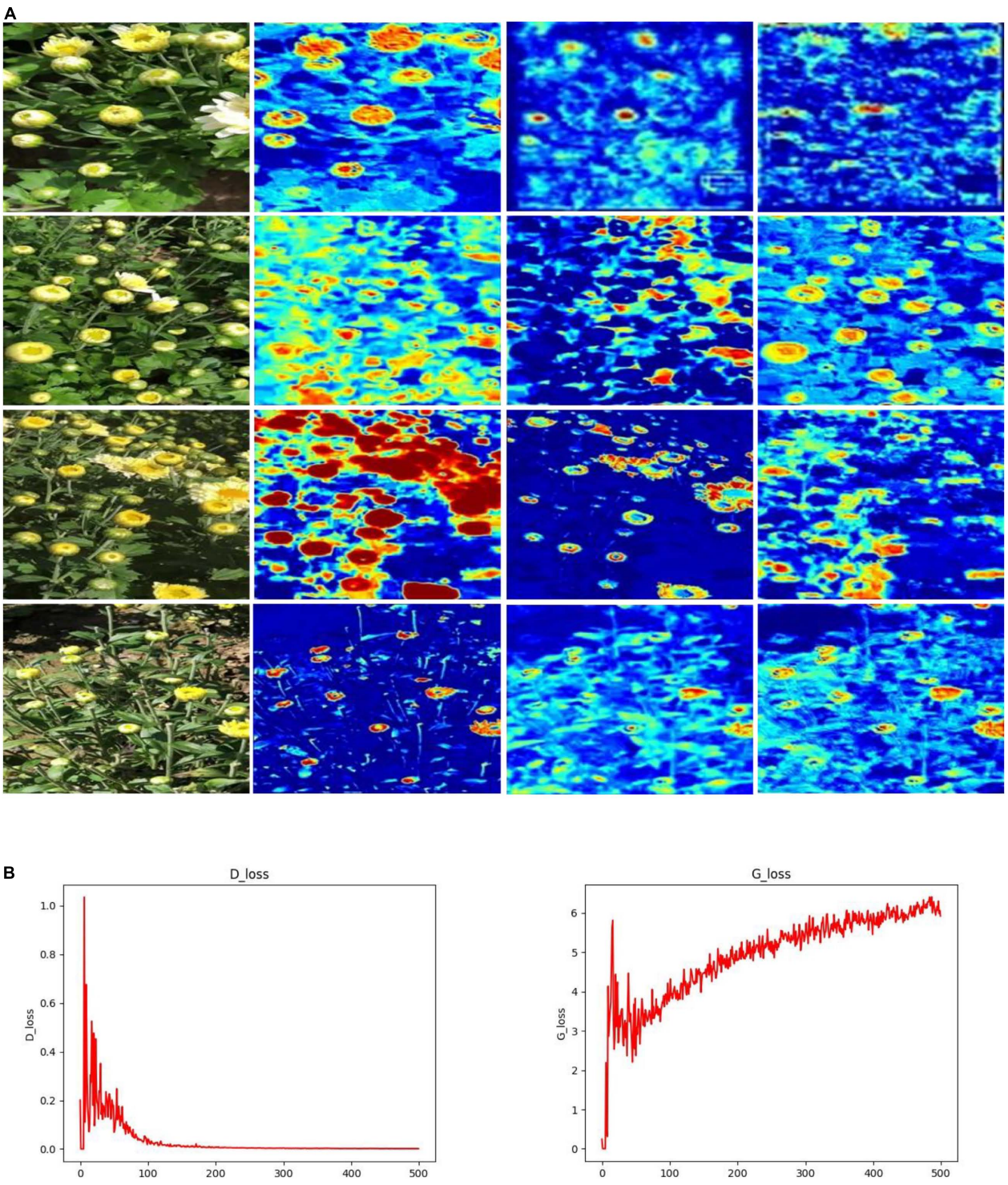


FIGURE 8 | (A) Visualization results and **(B,C)** training process.

DISCUSSION

To investigate the three issues summarized in the section "Introduction," we proposed the TC-YOLO and compared its results with the related work in **Table 2**. Our proposed TC-GAN generates high resolution images (512 × 512), and the *E* section of the experimental results shows that high resolution images can significantly enrich environmental features and thus improve the robustness of the model. GAN is prone to pattern collapse and gradient vanishing during the

training process, resulting in the lack of diversity in the generated image features (Wang et al., 2021). TC-GAN is able to generate images containing complex unstructured environments including illumination, overlap and occlusion to gain the benefit for detection under field environments, whereas most of synthetic images generated from other GANs listed in **Table 2** provide limited diversity and clear backgrounds. To intuitively view the image features through the generation process, we show the visualization process and training process in TC-YOLO (**Figure 8**). It can be seen that the important part (flower heads) of the plants is clearly activated and captured with the TC-YOLO.

There are several points to be optimized for TC-GAN despite its good detection performance. First, currently, there are no suitable metrics to evaluate synthetic images. FID is a widely recognized metric for evaluating synthetic images, but the FID metric is dedicated to evaluating several specific datasets and is not applicable to customized datasets. We can only evaluate the quality of synthetic images by their detection results in an object detection model. Therefore, establishing a standard set of evaluation metrics is an urgent issue to be addressed. Next, the training cost of TC-GAN is expensive. As can be seen from **Table 6**, the training of the whole model takes 1460 min under the 16 GB video memory of Tesla P100, and an ordinary device is difficult to train effectively. Thus, the light weight of TC-GAN is beneficial to the promotion of the technology. Besides, according to the experimental results in section "Impact of Different Unstructured Environments on the TC-YOLO" of the experimental results, TC-GAN can not fully construct images well for the illumination environmental setting. Note that the lack of efficient interaction between the generator network and the discriminator network leads to constant oscillation in the gradient and difficulty in convergence, as shown in **Figure 8B**. This is still a challenge without fully addressed in generative adversarial networks, and we suggest more attention should be paid to solve this challenge. Finally, our proposed model was deployed in NVIDIA Jetson TX2 with approximately 0.1 s per chrysanthemum inference time (the image size is 416 × 416). It is not real-time performance, and this deserves further optimization for network architecture such as network pruning and quantization.

CONCLUSION

This article presents a novel generative adversarial network architecture TC-GAN for generating tea chrysanthemum images under unstructured environments (illumination, overlap, occlusion). The TC-YOLO model is able to generate images with a resolution of 512 × 512 and achieves the AP of 90.09%, showing supreme results with other state-of-the-art generative adversarial networks. Finally, we deployed and tested the TC-YOLO model in the NVIDIA Jetson TX2 for robotic harvesting and solar insecticidal lamps systems development, achieving approximately 0.1 s per image (512 × 512). The proposed TC-GAN has the potential to be integrated into selective picking robots and solar insecticide lamp systems via the NVIDIA Jetson TX2 in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

CQ: conceptualization, methodology, software, writing – original draft, and writing – review and editing. JG: conceptualization and writing – review and editing. KC: supervision and writing – review and editing. LS: supervision, project administration, funding acquisition, and writing – review and editing. SP: writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Senior Foreign Expert Admission Scheme (G2021145009L), Modern Agricultural Equipment and Technology Demonstration and Promotion Project in Jiangsu Province (NJ2021-11), Lincoln Agri-Robotics as part of the Expanding Excellence in England (E3) Program, and National Key Research and Development Plan Project: Robotic Systems for Agriculture RS-Agri, (Grant No. 2019YFE0125200).

ACKNOWLEDGMENTS

Appreciations are given to the editor and reviewers of the Journal.

REFERENCES

Alsamhi, S. H., Almalki, F. A., Afghah, F., Hawbani, A., Shvetsov, A. V., Lee, B., et al. (2022). Drones' edge intelligence over smart environments in B5G: blockchain and federated learning synergy. *IEEE Trans. Green Commun. Netw.* 6, 295–312. doi: 10.1109/tgcn.2021.3132561

Alsamhi, S. H., Almalki, F. A., Al-Dois, H., Shvetsov, A. V., Ansari, M. S., Hawbani, A., et al. (2021). Multi-drone edge intelligence and SAR smart wearable devices for emergency communication. *Wirel. Commun. Mob. Com.* 2021:6710074.

Ansari, M. S., Alsamhi, S. H., Qiao, Y., Ye, Y., and Lee, B. (2020). "Security of distributed intelligence in edge computing: threats and countermeasures," in *The Cloud-to-Thing Continuum. Palgrave Studies in Digital Business & Enabling Technologies*, eds T. Lynn, J. Mooney, B. Lee, and P. Endo (Cham: Palgrave Macmillan), 95–122. doi: 10.1007/978-3-030-41110-7_6

Bian, Y., Wang, J., Jun, J. J., and Xie, X.-Q. (2019). Deep convolutional generative adversarial network (dcGAN) models for screening and design of small molecules targeting cannabinoid receptors. *Mol. Pharm.* 16, 4451–4460. doi: 10.1021/acs.molpharmaceut.9b00500

Cao, L., Zhang, X., Pu, J., Xu, S., Cai, X., and Li, Z. (2020). "The field wheat count based on the efficientdet algorithm," in *Proceedings of the 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, (New Delhi: ICISCAE). 557–561.

Chao, W., Wenhui, W., Jiahan, D., and Guangxin, G. (2021). "Research on network intrusion detection technology based on dcgan," in *Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, (Vienna: IAEAC). 1418–1422. doi: 10.3390/s19143075

Collier, E., Duffy, K., Ganguly, S., Madanguit, G., Kalia, S., Shreekant, G., et al. (2018). "Progressively growing generative adversarial networks for high resolution semantic segmentation of satellite images," in *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, (New Jersey, NJ: IEEE). 763–769.

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* 21:4749. doi: 10.3390/s21144749

Douarre, C., Crispim-Junior, C. F., Gelibert, A., Tougne, L., and Rousseau, D. (2019). Novel data augmentation strategies to boost supervised segmentation of plant disease. *Comput. Electron. Agric.* 165:104967. doi: 10.1016/j.compag.2019.104967

Espejo-Garcia, B., Mylonas, N., Athanasakos, L., Vali, E., and Fountas, S. (2021). Combining generative adversarial networks and agricultural transfer learning for weeds identification. *Biosyst. Eng.* 204, 79–89. doi: 10.1016/j.biosystemseng.2021.01.014

Gandhi, R., Nimbalkar, S., Yelamanchili, N., and Ponkshe, S. (2018). "Plant disease detection using CNNs and GANs as an augmentative approach," in *Proceedings of the 2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, (New Jersey, NJ: IEEE). 1–5.

Gong, X. Y., Chang, S. Y., Jiang, Y. F., and Wang, Z. Y. (2019). "AutoGAN: neural architecture search for generative adversarial networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), (Seoul, South Korea)*, (New Jersey, NJ: IEEE). 3223–3233.

He, Y., Zhang, J., Shan, H., and Wang, L. (2019). Multi-Task GANs for view-specific feature learning in gait recognition. *IEEE Trans. Inf. Forensic Secur.* 14, 102–113. doi: 10.1109/tifs.2018.2844819

Hu, G., Wu, H., Zhang, Y., and Wan, M. (2019). A low shot learning method for tea leaf's disease identification. *Comput. Electron. Agric.* 163:104852. doi: 10.1016/j.compag.2019.104852

Hu, G., Yin, C., Wan, M., Zhang, Y., and Fang, Y. (2020). Recognition of diseased pinus trees in UAV images using deep learning and AdaBoost classifier. *Biosyst. Eng.* 194, 138–151. doi: 10.1016/j.biosystemseng.2020.03.021

Jeon, H., and Lee, D. (2021). A New Data augmentation method for time series wearable sensor data using a learning mode switching-based DCGAN. *IEEE Robot. Autom. Lett.* 6, 8671–8677. doi: 10.1109/lra.2021.3103648

Kim, C., Park, S., and Hwang, H. J. (2021). Local stability of wasserstein GANs with abstract gradient penalty. *IEEE Trans. Neural Netw. Learn. Syst.* 1–11. doi: 10.1109/TNNLS.2021.3057885

Kim, S. W., Kook, H. K., Sun, J. Y., Kang, M. C., and Ko, S. J. (2018). "Parallel feature pyramid network for object detection," in *Proceedings of the European Conference on Computer Vision (ECCV), (Munich, Germany)*, (Tel Aviv: ECCV). 239–256.

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.* 21:5386. doi: 10.3390/s21165386

Liu, J., and Wang, X. W. (2020). Tomato diseases and pests detection based on improved yolo V3 convolutional neural network. *Front. Plant Sci.* 11:898. doi: 10.3389/fpls.2020.00898

Liu, L., Zhang, H., Xu, X., Zhang, Z., and Yan, S. (2020). Collocating clothes with generative adversarial networks cosupervised by categories and attributes: a multidiscriminator framework. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 3540–3554. doi: 10.1109/TNNLS.2019.2944979

Liu, X., Chen, S., Song, L., Woźniak, M., and Liu, S. (2021). Self-attention negative feedback network for real-time image super-resolution. *J. King Saud. Univ-Com. Info. Sci.* doi: 10.1016/j.jksuci.2021.07.014

Liu, X., and Hsieh, C. (2019). "Rob-GAN: generator, discriminator, and adversarial attacker," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (New Jersy, NJ: IEEE). 11226–11235.

Liu, Z. L., Wang, J., Tian, Y., and Dai, S. L. (2019). Deep learning for image-based large-flowered chrysanthemum cultivar recognition. *Plant Methods* 15:146. doi: 10.1186/s13007-019-0532-7

Luo, Z., Yu, H., and Zhang, Y. (2020). Pine cone detection using boundary equilibrium generative adversarial networks and improved YOLOv3 model. *Sensors* 20:4430. doi: 10.3390/s20164430

Mao, X., Li, Q., Xie, H., Lau, R. Y. K., Wang, Z., and Smolley, S. P. (2019). On the effectiveness of least squares generative adversarial networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 2947–2960. doi: 10.1109/TPAMI.2018.2872043

Marino, S., Beauseroy, P., and Smolarz, A. (2020). Unsupervised adversarial deep domain adaptation method for potato defects classification. *Comput. Electron. Agric.* 174:105501. doi: 10.1016/j.compag.2020.105501

Mufti, A., Antonelli, B., and Monello, J. (2019). "Conditional gans for painting generation," in *Proceedings of the 12th International Conference on Machine Vision (ICMV), (Amsterdam, Netherlands)*, (Boulevard Victor Hugo: ICMV). 1143335.

Nafi, N. M., and Hsu, W. H. (2020). "Addressing class imbalance in image-based plant disease detection: deep generative vs. sampling-based approaches," in *Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, (Bratislava: IWSSIP). 243–248.

Nazki, H., Yoon, S., Fuentes, A., and Park, D. S. (2020). Unsupervised image translation using adversarial networks for improved plant disease recognition. *Comput. Electron. Agric.* 168:105117. doi: 10.1016/j.compag.2019.105117

Olatunji, J. R., Redding, G. P., Rowe, C. L., and East, A. R. (2020). Reconstruction of kiwifruit fruit geometry using a CGAN trained on a synthetic dataset. *Comput. Electron. Agric.* 177:12.

Padilla-Medina, J. A., Contreras-Medina, L. M., Gavilán, M. U., Millan-Almaraz, J. R., and Alvaro, J. E. (2019). Sensors in precision agriculture for the monitoring of plant development and improvement of food production. *J. Sens.* 2019:7138720.

Qi, C., Gao, J., Pearson, S., Harman, H., Chen, K., and Shu, L. (2022). Tea chrysanthemum detection under unstructured environments using the TC-YOLO model. *Expert Syst. Appl.* 193:116473. doi: 10.1016/j.eswa.2021.116473

Qiao, K., Chen, J., Wang, L. Y., Zhang, C., Tong, L., and Yan, B. (2020). BigGAN-based bayesian reconstruction of natural images from human brain activity. *Neuroscience.* 444, 92–105. doi: 10.1016/j.neuroscience.2020.07.040

Shete, S., Srinivasan, S., and Gonsalves, T. A. (2020). TasselGAN: an application of the generative adversarial model for creating field-based maize tassel data. *Plant Phenomics* 2020:8309605. doi: 10.34133/2020/8309605

Talukdar, B. (2020). "Handling of class imbalance for plant disease classification with variants of GANs," in *Proceedings of the 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, (Sutton: ICIIS). 466–471.

Tran, N. T., Bui, A., and Cheung, N. M. (2018). "Dist-GAN: an improved GAN using distance constraints," in *Proceedings of the European Conference on Computer Vision (ECCV), (Munich, Germany)*, (Colorado, CO: ECCV). 387–401. doi: 10.1007/978-3-030-01264-9_23

Tran, N. T., Tran, V. H., Nguyen, N. B., Nguyen, T. K., and Cheung, N. M. (2021). On Data Augmentation for GAN Training. *IEEE Trans. Image Process.* 30, 1882–1897. doi: 10.1109/TIP.2021.3049346

Wang, C., Xu, C., Yao, X., and Tao, D. (2019). Evolutionary generative adversarial networks. *IEEE Trans. Evol. Comput.* 23, 921–934.

Wang, Z., She, Q., and Ward, T. E. (2021). Generative adversarial networks in computer vision: a survey and taxonomy. *ACM Comput. Surv.* 37, 1–38. doi: 10.1145/3439723

Wieczorek, M., Sika, J., Wozniak, M., Garg, S., and Hassan, M. (2021). Lightweight CNN model for human face detection in risk situations. *IEEE Trans. Ind. Info.* 1–1. doi: 10.1109/tii.2021.3129629

Yang, L., Fan, W., and Bouguila, N. (2020). Clustering analysis via deep generative models with mixture models. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 340–350. doi: 10.1109/TNNLS.2020.3027761

Yue, J., Zhu, C., Zhou, Y., Niu, X., Miao, M., Tang, X., et al. (2018). Transcriptome analysis of differentially expressed unigenes involved in flavonoid biosynthesis during flower development of Chrysanthemum morifolium 'Chuju'. *Sci. Rep.* 8:13414. doi: 10.1038/s41598-018-31831-6

Zhang, T., and Li, L. (2020). "An improved object detection algorithm based on M2Det," in *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, (New Delhi: ICAICA). 582–585.

Zhang, Z., Qiao, S., Xie, C., Shen, W., Wang, B., and Yuille, A. L. (2018). "Single-shot object detection with enriched semantics," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (New Jersy, NJ: IEEE). 5813–5821. doi: 10.1016/j.compbiomed.2020.103867

Zhao, J., Xiong, L., Li, J., Xing, J., Yan, S., and Feng, J. (2019). 3D-Aided Dual-agent GANs for unconstrained face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 2380–2394. doi: 10.1109/TPAMI.2018.2858819

Zhao, W., Yamada, W., Li, T. X., Digman, M., and Runge, T. (2021a). Augmenting crop detection for precision agriculture with deep visual transfer learning-a case study of bale detection. *Remote Sens.* 13:17.

Zhao, Y., Chen, Z., Gao, X., Song, W., Xiong, Q., Hu, J., et al. (2021b). Plant disease detection using generated leaves based on doubleGAN. *IEEE-ACM Trans. Comput. Biol. Bioinfo.* 1–1. doi: 10.1109/TCBB.2021.3056683

Zheng, J. Y., Lu, B. Y., and Xu, B. J. (2021). An update on the health benefits promoted by edible flowers and involved mechanisms. *Food Chem.* 340:17. doi: 10.1016/j.foodchem.2020.127940

Zhong, F., Chen, Z., Zhang, Y., and Xia, F. (2020). Zero- and few-shot learning for diseases recognition of Citrus aurantium L. using conditional adversarial autoencoders. *Comput. Electron. Agric.* 179:105828. doi: 10.1016/j.compag.2020.105828

Zhou, P., Gao, B., Wang, S., and Chai, T. (2022). Identification of abnormal conditions for fused magnesium melting process based on deep learning and multisource information fusion. *IEEE T. Ind. Electron.* 69, 3017–3026. doi: 10.1109/tie.2021.3070512

Zhou, X., Liang, W., Shimizu, S., Ma, J., and Jin, Q. (2021). Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems. *IEEE Trans. Ind. Info.* 17, 5790–5798. doi: 10.1109/tii.2020.3047675

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fpls.2022.852375

Measurement and Analysis of Root Anchorage Effect on Stalk Forces in Lodged Corn Harvesting

Qiankun Fu[1,2], Jun Fu[1,2], Zhi Chen[1,3], Chao Chen[1,2], Jialiang Zhang[1,2] and Luquan Ren[1,2]*

[1] College of Biological and Agricultural Engineering, Jilin University, Changchun, China, [2] Key Laboratory of Bionics Engineering, Ministry of Education, Jilin University, Changchun, China, [3] Chinese Academy of Agricultural Mechanization Sciences, Beijing, China

The effect of root anchorage on corn stalk is the main cause of difficulties in stalk lifting and ear picking of lodged corn. To quantify the forces on the stalks caused by root anchorage in corn harvesting, a root force measurement system was designed and applied in this study. The bending moment and torsional moment on the upright and lodged corn stalks were measured in corn harvesting with the designed system and the results were compared with the manually measured failure boundaries. The manually measured results showed bending moments to push down the upright stalks, to lift the lodged corn stalks, and to slip the lodged corn stalks were 35.12, 23.33, and 40.36 Nm, respectively, whereas the torsional moments needed to twist off the upright and lodged corn stalks were 4.02 and 3.33 Nm, respectively. The bending moments that the corn header applied to the upright, forward lodged, reverse lodged, and lateral lodged corn stalks were 10.68, 22.24, 16.56, and 20.42 Nm, respectively, whereas the torsional moments on them were 1.32, 1.59, 1.55, and 1.77 Nm, respectively. The bending force was the main factor that broke the root anchorage and influenced the stalk movement of lodged corn in harvesting. By analyzing the bending moment curves on the lodged corn stalks, it was proposed that for the harvesting of corn lodged in the forward, reverse, and lateral direction, the corresponding harvester header improvement suggestions are enlarging the size of pins on the gathering chains, reducing the speed of gathering chains, and lengthening the snouts with a sleeker surface, respectively. This study provides base data for the root anchorage effect on lodged corn and provides references for the improved design of the corn harvester header.

Keywords: root anchorage, lodged corn, corn harvest, harvester header, bending moment, static torque sensor

OPEN ACCESS

Edited by:
Daobilige Su,
China Agricultural University, China

Reviewed by:
Wenshuai Liu,
Shaanxi University of Technology, China
Hu Hongnan,
Zhongkai University of Agriculture and Engineering, China

***Correspondence:**
Jun Fu
fu_jun@jlu.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *11 January 2022*
Accepted: *22 February 2022*
Published: *12 April 2022*

Citation:
Fu Q, Fu J, Chen Z, Chen C, Zhang J and Ren L (2022) Measurement and Analysis of Root Anchorage Effect on Stalk Forces in Lodged Corn Harvesting. Front. Plant Sci. 13:852375. doi: 10.3389/fpls.2022.852375

INTRODUCTION

Corn lodging is usually caused by excessive planting density, improper use of fertilizer, unreasonable irrigation or diseases, and pests during the growth period (Echezona, 2007; Shelby et al., 2018; Liu et al., 2021). The bending strength and rind puncture resistance of stalks are usually taken as the indices for lodging resistance in variety breeding (Albrecht et al., 1986; Robertson et al., 2016; Seegmiller et al., 2020). The main effect of lodging is to reduce crop yield (Ma et al., 2014). The lodging area statistics and yield reduction prediction are usually carried out with satellite remote

sensing and UAV images (Han et al., 2018; Chauhana et al., 2019; Wilke et al., 2019). Lodging in the corn mature period such as the late milky ripeness stage and the wax ripeness stage is usually caused by extreme weather such as heavy rains and rainstorms. This kind of lodging is characterized as whole plant inclination because of the loosening of root-soil (Martinez-Vazquez, 2016). Lodging in the corn maturation period causes severe ear loss in harvesting (Wang et al., 2021) because the harvester would not be able to pick corn ears lower than the working height of the corn header (Paulsen et al., 2014; Xue et al., 2020a). Sugarcane has the same lodging morphology as corn stalk. The lodged sugarcane is lifted with the spiral dividers in the harvest (Bai et al., 2021). Short-stem crops such as wheat and rice are harvested with lowered headers to cut off the crop and feed the whole plant to the threshing part (Paulsen et al., 2014; Phetmanyseng et al., 2019).

Different from the harvest of sugarcane and the short-stem crops, the stalk is not cut in corn harvesting. Lowering header height and applying headers with narrow row spacing units are the main compromise means in the lodged corn harvesting (Yang et al., 2016; Wang et al., 2021). In the harvesting, the snouts of the header are extended into the bottom of the stalks to lift them with the travel of the harvester. The stalks are fed into the gap between the snapping plates and pulled down by the stalk rolls under the snapping plates. Then the corn ears are picked by the blocking effect of the snapping plates, as shown in **Figure 1**. In this process, the corn stalk is not only subjected to forces by the stalk rolls and the pins on the gathering chains but also influenced by the root anchorage (Fu et al., 2019). In the previous studies, Donovan et al. (1982) measured the pulling force of the corn root with a sensor connected to the three-point linkage of the tractor. Reneau et al. (2020) tested the contribution of brace roots on anchorage by measuring the deflection forces after removing them. However, there is still a lack of reliable data about the root anchorage effect on the lodged corn stalks in corn harvesting. The technical difficulties such as stalk lifting, stalk feeding, and header blockage could not be solved without reliable analysis. It is also impossible to make a reasonably improved design on corn harvester header in dealing with lodged corn.

Bending failure and torsional failure are the main forms of stalk failure in lodged corn harvesting. The bending moment and torsional moment on the stalks could be taken as the indices for the root anchorage in corn harvesting (Francisco et al., 2018; Stubbs et al., 2019). This study focused on the forces acting on upright and lodged corn stalks affected by the root anchorage in the lodged corn harvesting. It starts with the measurement of the bending moment and torsional moment to cause stalk failure on the upright and lodged corn stalks. Then the bending moment and torsional moment on the upright and lodged corn stalks exerted by the corn header were measured in the field test with a designed measurement system. The force analysis was made on the stalks by comparing the test results with the failure boundaries. The action laws of corn header on the upright stalks, forward lodged corn stalks, reverse lodged corn stalks, and lateral lodged corn stalks were obtained with the analysis of bending moment curves on corn stalks. The causes of ear miss picking and header blockage were discussed, and suggestions for corn header improved design in lodged corn harvest were propounded. The results of this study will give base data for the root anchorage effect on lodged corn and provide references for the improved design of the corn harvester header.

MATERIALS AND METHODS

Mechanical Properties of Corn Stalks

The mechanical properties of corn stalk are the physical basis for the stress analysis of stalk in corn harvesting. The parameters measured in this test were the force required to push down the upright stalks, the lifting force of the lodged corn stalks, the slipping force of the lodged corn stalks, and the torsional moments to twist off of upright and lodged corn stalks. The tested variety in this study was Xianyu 335, which was widely planted in Northeast China. The test was made in Changpaozi Village, Yitong Manchu Autonomous County, Siping City, Jilin Province. The corn lodged in 3 typhoons between Aug 27 and Sep 8, 2020, about 30–40 days before the test.

Pushing Down Force of the Upright Corn

A hand-held dynamometer was used to measure the forces. The dynamometer was an HP-300 model with a relative error of 0.5%, which was manufactured by Yueqing Handpi Instruments Co., Ltd. A horizontal force was applied to the upright corn stalks with the dynamometer until the corn was pushed down to the ground. This maximum force value in pushing down the upright corn stalk was recorded. The stalk failure might occur at the root of corn when the soil loosened or at any position below the force application position when the stalk broke. The measurement method is shown in **Figure 2A**. To compare the measured force with the forces in corn harvesting, the force F_{bend} was applied at the height of 400 mm above the ground, which was approximately equal to the minimum working height of the corn header according to the harvester operating instructions. The height was approximate to the force application height of the corn header to corn stalks. So were the heights $H_{measure}$ in the measurement of F_{lift} and F_{slip} in the following tests. The bending moments on the stalks were obtained with the product of the force and the height.

Lifting Force of Lodged Stalk

In the harvest of lodged corn, the snout tips of the corn header extend into the stalks' bottom, the pins on the gathering chains lift the stalks with corn ear to the corn header's working height. Therefore, the stalk lifting force of lodged corn stalks was measured. The measurement method is shown as F_{lift} in **Figure 2B**. The maximum force was recorded until the ear was lifted to the height of 500 mm. This height was determined according to the working height of the applied corn harvester header, which was 400 mm. The bending moments were obtained with the product of F_{lift} and $H_{measure}$.

The Slipping Force of Lodged Corn Stalks

When the lodged corn stalk is subjected to a lateral force, it may slide on the ground. Since the corn root is fixed by the soil, and the top of the corn stalk can be regarded as a free end, when

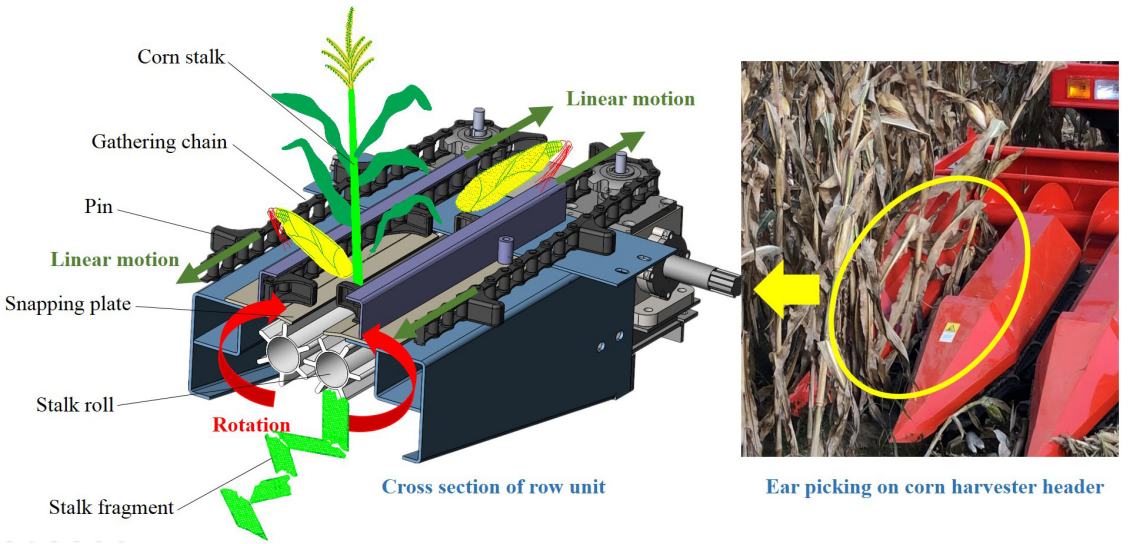


FIGURE 1 | Process of ear picking on the corn header.

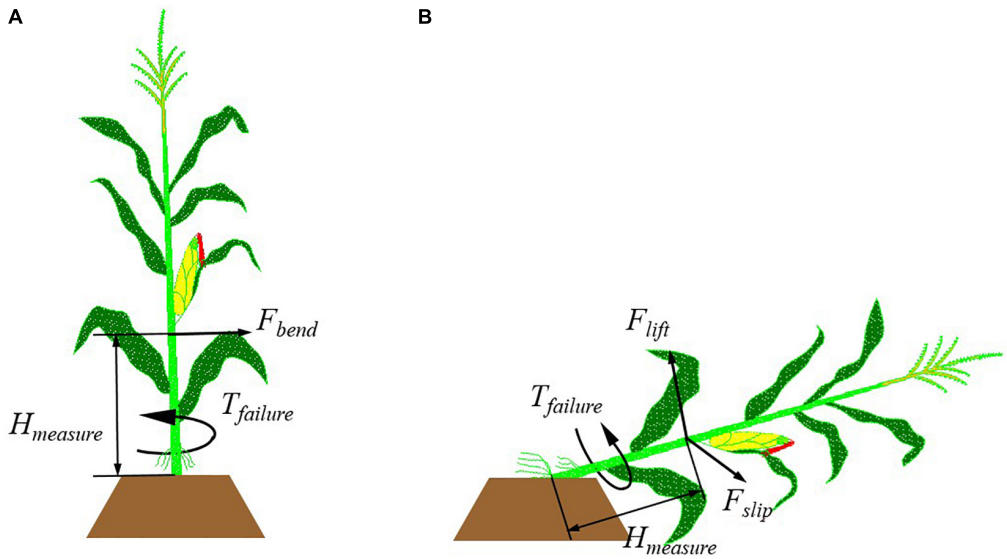


FIGURE 2 | Measurement method of the mechanical properties of upright and lodged corn stalk. **(A)** F_{bend} and $T_{failure}$ on the upright stalks; **(B)** F_{lift}, F_{slip}, and $T_{failure}$ on the lodged corn stalks.

the lodged corn stalk is subjected to a horizontal force, the corn stalk would slide around the root on the ground. This sliding will cause the stacking of stalks in front of the corn header and cause corn header blockage eventually. It should be avoided in corn harvesting. Therefore, the critical condition for the occurrence of stalk slipping was measured. The measurement method of this force is shown as F_{slip} in **Figure 2B**.

Torsional Moment Boundaries of Corn Stalk Failure

Torsion is also considered an important reason for stalk failure (Faisal et al., 2017). In this test, a torsional force was applied to the upright and lodged corn stalks to get the failure boundary of the stalks under the torsional moment. The torsional moment was recorded as $T_{measure}$ when the stalk was twisted off or the root is ripped out from the soil. The measurement method on the upright and lodged corn stalks is shown in **Figures 2A,B**. The torsional moment was applied and measured with a torque wrench (PLARZ-30 Nm by Suzhou Duotong Hardware Electrical Co., Ltd.). The stalks were clamped with a fixture on the top of the torque wrench. The abrasive cloth was wrapped around the corn stalk to increase friction force because the friction between the stalk and the steel clamping fixture was too small to prevent relative sliding. When the stalks got lodged, the torsional forces could be applied to any position of the stalks by the

FIGURE 3 | An experiment of forces on stalks in corn harvesting. **(A)** Experiment on the upright corn; **(B)** experiment on the lodged corn; **(C)** the experiment scene of corn harvest.

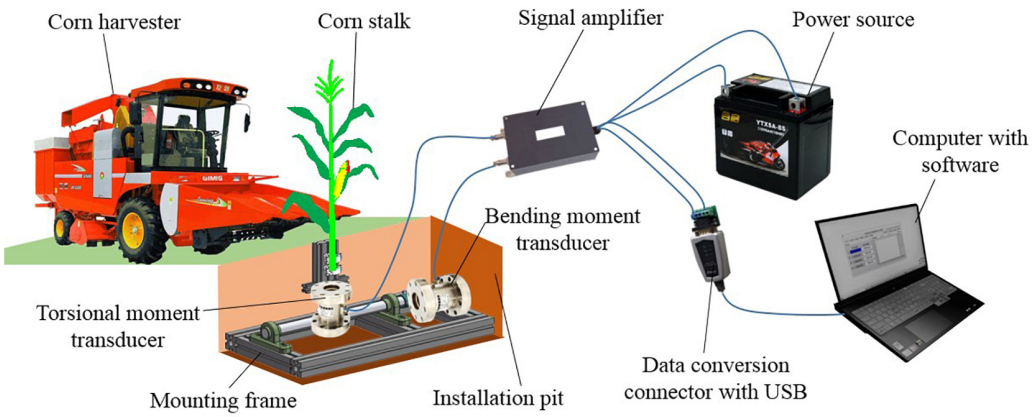


FIGURE 4 | Measurement of forces on corn stalks in harvesting.

harvester header snouts. To reduce the interference of stalk elastic deformation on the measurement results, and avoid the influence of brace roots on stalk clamping, the measurement was made at the height of 150 mm above the ground.

Forces on the Stalks in Corn Harvesting

In this section, the torsional moment and bending moment on the upright and lodged corn stalks in harvesting were measured. According to the relationship between corn lodging direction and harvester travel direction, the lodged corn stalks were classified into forward, reverse, and lateral, corresponding to the harvester traveling in the same, opposite, and vertical direction with stalk lodging. The harvester applied in the experiment was the 4YZP-4Y corn harvester manufactured by Juming Company in Shandong, China. The speed of the harvester was reduced to 0.5 m/s to meet the operation requirements of lodged corn, as its standard working speed was 0.55–1.1 m/s according to the operating instructions. In the measurement, the measured corn plant was fixed on the mounting frame. The mounting frame was placed in an installation pit. It ensured that the measured corn plant was at the same height with natural growth and avoided the collision between the harvester and the mounting frame in harvesting. **Figure 3** shows the scene of the field experiment.

Measurement System

Figure 4 shows the measurement system for the forces on the stalk in corn harvesting. The system consisted of a corn harvester, an installation pit to place the mounting frame, a mounting frame with a torsional moment transducer and bending moment transducer, a whole corn plant fixed on the mounting frame, a signal amplifier, a data conversion connector with USB, a computer with acquisition software, and a 12 V power source. The power of the sensors and the signal amplifier were supplied by a 12 V power source. The data were displayed on the acquisition software of the computer after format transformation.

The structure of the mounting frame is shown in **Figure 5**. It was assembled from profiles. The upper part of the mounting frame was a stalk clamp to fix the stalk. At the bottom of the stalk clamp was the torsional moment transducer (ZNNF-5 Nm static torque sensor by Bengbu Zhongnuo Sensor Co., Ltd.). When the stalk was twisted, the torque signal would be collected by the torsional moment transducer. The torsional moment

transducer was fixed on the horizontal shaft with U-bolts. It could rotate around the horizontal shaft to simulate the morphology of the upright and lodged corn stalks at different angles. The horizontal shaft was installed on the frame through bearings with the seat. The bending moment transducer (ZNNF-100 Nm static torque sensor by Bengbu Zhongnuo Sensor Co., Ltd.) was connected with the horizontal shaft at one end and fixed with the frame at the other end. When the stalk was subjected to forces by the harvester, the bending moments would be collected in the form of static torque. There were four fixing wedges at the bottom of the frame. They could be stuck into the soil to prevent frame movement.

Measurement Method

The tests were conducted on upright and lodged corn. The bending moments in the harvester traveling direction and the torsional moments were measured on the upright corn, as shown in **Figure 6A**. In the harvest of forward and reverse lodged corn, the bending moments in the harvester traveling direction and the torsional moments on the stalks were collected, as shown in **Figures 6B,C**. When the harvesting was lateral to the corn lodging direction as **Figure 6D**, the bending moments in the stalk lifting direction and the torsional moments on the stalks were measured. Each measurement was repeated 60 times.

RESULTS AND DISCUSSION

Properties of the Upright and Lodged Corn Stalks

Force Bearing Capacities

Figure 7 shows the maximum forces that the upright and lodged corn stalks could bear under different force conditions. The average force to push down the upright corn stalks was 87.80 ± 29.89 N, the average lifting force of the lodged corn stalks was 58.33 ± 10.76 N, and the average slipping force of lodged corn stalks was 100.91 ± 28.79 N. The bending moments on the stalks under the three situations were 35.12, 23.33, and 40.36 Nm, respectively, supposing the effects of the adjacent stalks not considered. The failure bending moment of the upright stalks was larger than Sekhon et al. (2020) measured with the equipment of DARLING on corn 40 days after anthesis, but it was close to Xue et al. (2020b)'s result measured at the ear position after physical maturity. The difference might result from the varieties, growth stages, and the stalk conditions caused by planting patterns. The upright stalks usually broke at the bottom internodes. It showed that the anchoring effect of the root was strong, just as Donovan et al. (1982) verified in the test. It was measured that the pulling force of the root varied between 836 and 1767 N. However, it was much easier to lift the lodged corn stalks than push down the upright stalks. Because the soil structure was destroyed in lodging and the fixation of soil on the root decreased greatly in the lodging direction. The force to slip the lodged corn stalks was much larger. It could be ascribed to the anchorage of root in the slipping direction, even though it was weakened in the lifting direction. This phenomenon was proved in the anchorage of brace roots by Reneau et al. (2020). The force to slip lodged stalk was a little larger than that to push down the upright stalk because the stacking of lodged stalks restricts the slipping.

Failure Torsional Moment of Corn Stalks

The torsional moment required to twist the upright and lodged corn stalks is shown in **Figure 8**. The average failure torsional moment of the upright stalks was 4.02 ± 0.84 Nm, and that of lodged corn stalks was 3.33 ± 0.92 Nm. The weaker torsional bearing capacity of lodged corn stalks might be attributed to

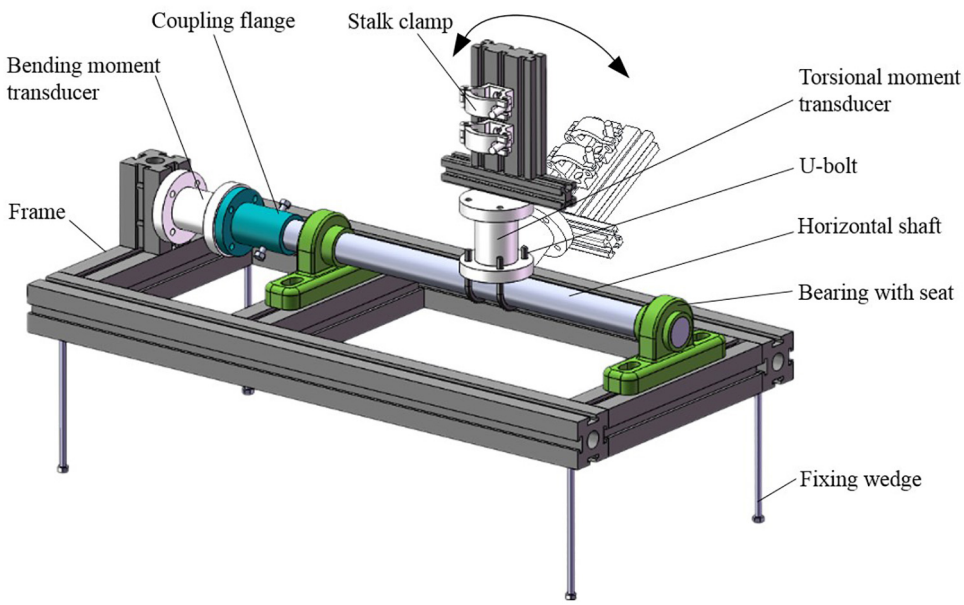


FIGURE 5 | Structural diagram of the mounting frame.

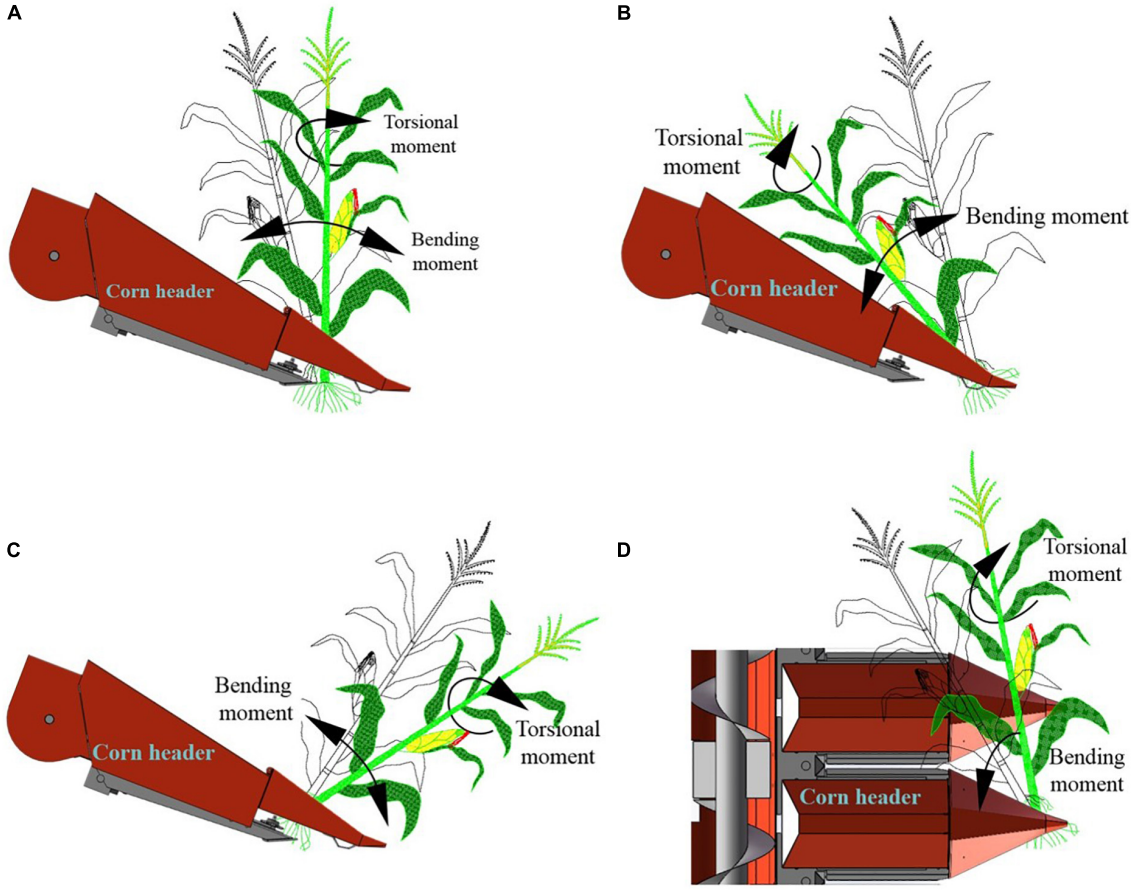


FIGURE 6 | Measurement methods of forces on upright and lodged corn stalks. **(A)** Upright stalks; **(B)** forward lodged stalks; **(C)** reverse lodged stalks; **(D)** lateral lodged stalks.

the decreasing of moisture content because of the insufficient supply of water after lodging, just like the declining of stalk strength after corn maturity (Xue et al., 2020b). Additionally, the destruction of soil also made the root easier to rotate under torsional force. Windsor-Collins et al. (2007) and Faisal et al. (2017) had measured the torsional force of palm petals and plant petioles similarly in the previous studies. But a few data were recorded on the torsional moment of corn stalks, which may be because the clamping devices could not apply sufficient friction on the sleek surface of the corn stalk. The problem was solved with the wound abrasive cloth between the clamping device and the corn stalks in this experiment.

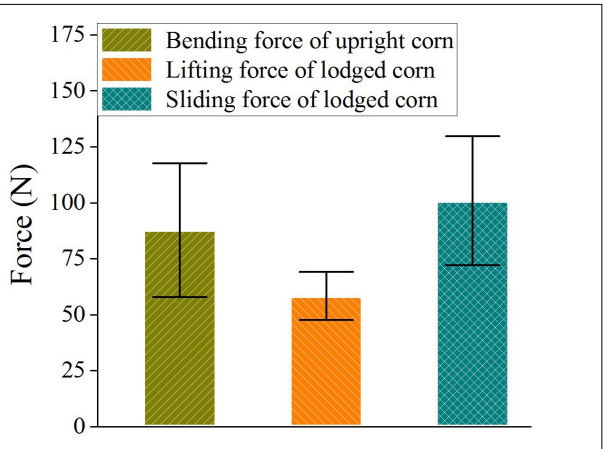


FIGURE 7 | Force-bearing capacities of upright and lodged corn stalks. Error bars indicate standard deviations.

Effect of Root Anchorage on the Stalk in Corn Harvesting

Bending Moment and Torsional Moment on Corn Stalk

The maximum and minimum values of bending moments and torsional moments on the stalks in corn harvesting are shown in **Table 1**, and the distributions of maximums are plotted in **Figure 9**. For the bending moment on the upright stalks and the torsional moments on all stalks, the direction in which the maximum bending moments appeared was specified as positive. But for the lodged corn stalks, the stalk lifting direction was specified as positive.

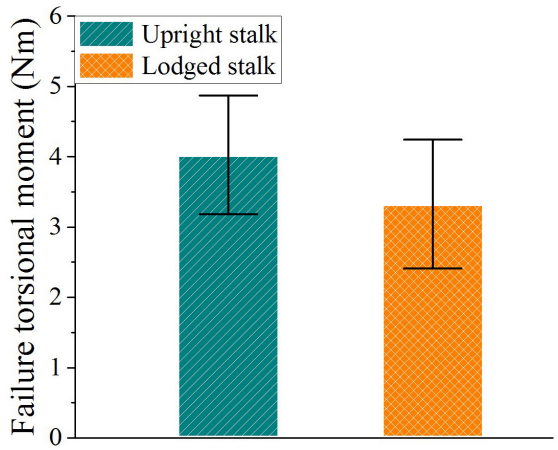


FIGURE 8 | Failure torsional moment of upright and lodged corn stalks. Error bars indicate standard deviations.

The bending moments on the lodged corn stalks were greater than that on the upright stalks, which indicated larger forces on the lodged corn. If the bending force was in the same direction with corn lodging, it would push the stalks to the ground. If it was opposite to the lodging direction, it would lift the stalks.

The maximum bending moments and torsional moments on the upright stalks were approximately 30% to the stalks' failure boundaries. **Figure 9A** shows that the bending moments of the upright stalks varied in a small range. The speed calculations in corn header design ensured that the upright stalks would not be pushed down by the harvester header in harvesting (Wang et al., 2016).

When the harvester traveled in the same direction with corn lodging, the stalks were forced backward by the pins on the gathering chains of the corn header. **Table 1** shows that the average bending moment on the stalks was 22.24 Nm, which was near to the bending moment calculated with the lifting force of the lodged corn stalks. When the harvester traveled opposite to the corn lodging direction, the average bending moment on the corn stalks was 16.56 Nm, which was at approximately two-thirds to the lifting moment of the lodged corn stalks. **Figure 9A** indicates many of the tested stalks were not lifted. When the harvester traveled lateral to the corn lodging direction, the average bending moment on the stalks was 20.42 Nm. It can be inferred that most of the stalks were lifted when the header snouts extended into the stalk bottom.

It can be seen from **Figure 9B** that the bending moments on lodged corn stalks did not show a definitive increase compared to the upright stalks. These bending moments were around half of the stalk failure torsional moment and could not fracture the stalks in corn harvesting. But the stalks were more susceptible to failure when the bending load was combined with torsional moment (Stubbs et al., 2019).

Bending Moment Curves Analysis

The bending was the main cause of stalk failure in corn harvesting. The typical bending moment curves on the stalks under the tested working conditions are shown in **Figure 10**.

In the harvest of upright corn (**Figure 10A**), the ear picking was completed in a short time. The bending moments reached the maximum shortly after the harvest began. For the forward lodged corn stalks (**Figure 10B**), the bending moment peak values were larger than that of the upright corn. The bending moments reached the maximum in a short time and declined with fluctuations with the ear picking. Contrarily, in the harvesting of reverse lodged corn stalks, the bending moments were small at the beginning and got larger gradually with fluctuations (**Figure 10C**). **Figure 10D** shows the bending moments on the lateral lodged corn stalks in harvesting. The stalks were lifted by the inclined surface of snouts with harvester traveling. The bending moments increased gradually at the beginning and fluctuated later.

The bending moment curves on the harvested corn stalks showed that the corn header has different operation characteristics for corn in upright and different lodging states. The operation characteristics depended on the structural and kinematic parameters of the corn header. The harvesting capacity of the corn header on lodged corn can be improved by optimizing the parameters of the harvester header.

Inspirations From the Bending Moment Curves

The contrast between the bending moment curves on the forward and reverse lodged corn stalks reflected the differences in force applied by the corn header. For the forward lodged corn, the forces were first applied to the lower part of a stalk, then the whole stalk was lifted by the pins on the gathering chain, as **Figure 11A**. The stalk was pulled upright in a short time. The stalk pulling and ear picking could be completed smoothly (**Figure 11B**). To improve the efficiency of stalk lifting and avoid ear loss caused by miss-lifting, the size of pins on the gathering chains should be enlarged. And the pins on both sides had better be arranged symmetrically rather than misplaced.

Forces on the reverse lodged corn were applied from the top of the stalks. Due to the stalk deformation and the anchorage of the root, the stalks could not be lifted and fed into the feeding space of the corn header rapidly as shown in **Figure 11C**. The

TABLE 1 | Bending moments and torsional moments on stalks in corn harvesting (Nm).

	Maximum bending moment	Minimum bending moment	Maximum torsional moment	Minimum torsional moment
Upright stalks	10.68 ± 2.35	−3.01 ± 2.87	1.32 ± 0.61	−0.26 ± 0.15
Forward lodged stalks	22.24 ± 5.27	−6.25 ± 4.16	1.59 ± 0.56	−0.37 ± 0.20
Reverse lodged stalks	16.56 ± 4.75	−4.82 ± 4.30	1.55 ± 0.55	−0.32 ± 0.25
Lateral lodged stalks	20.42 ± 4.96	−4.40 ± 3.65	1.77 ± 0.74	−0.50 ± 0.39

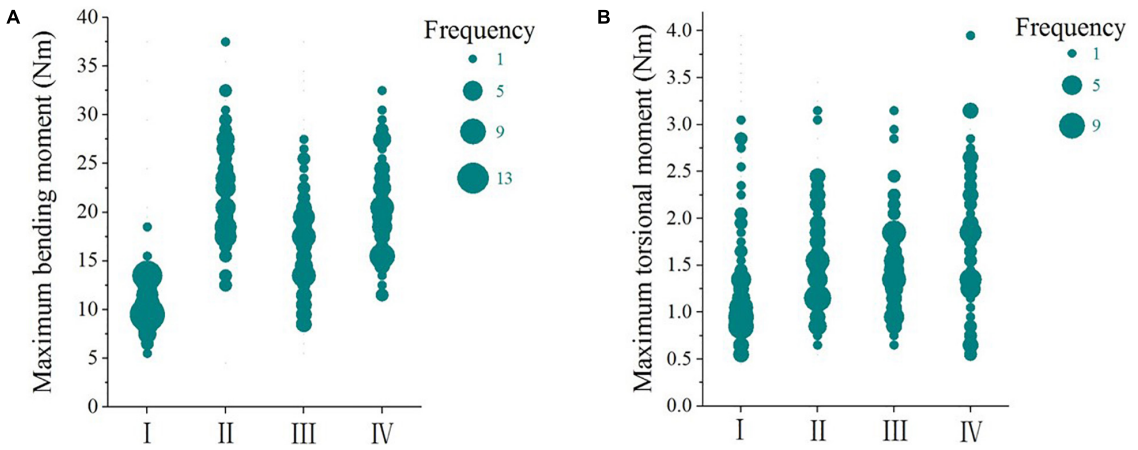


FIGURE 9 | Frequency distribution of bending moments and torsional moments on stalks in harvesting. **(A)** Bending moments; **(B)**, torsional moments; I upright stalks; II forward lodged stalks; III, reverse lodged stalks; IV, lateral lodged stalks.

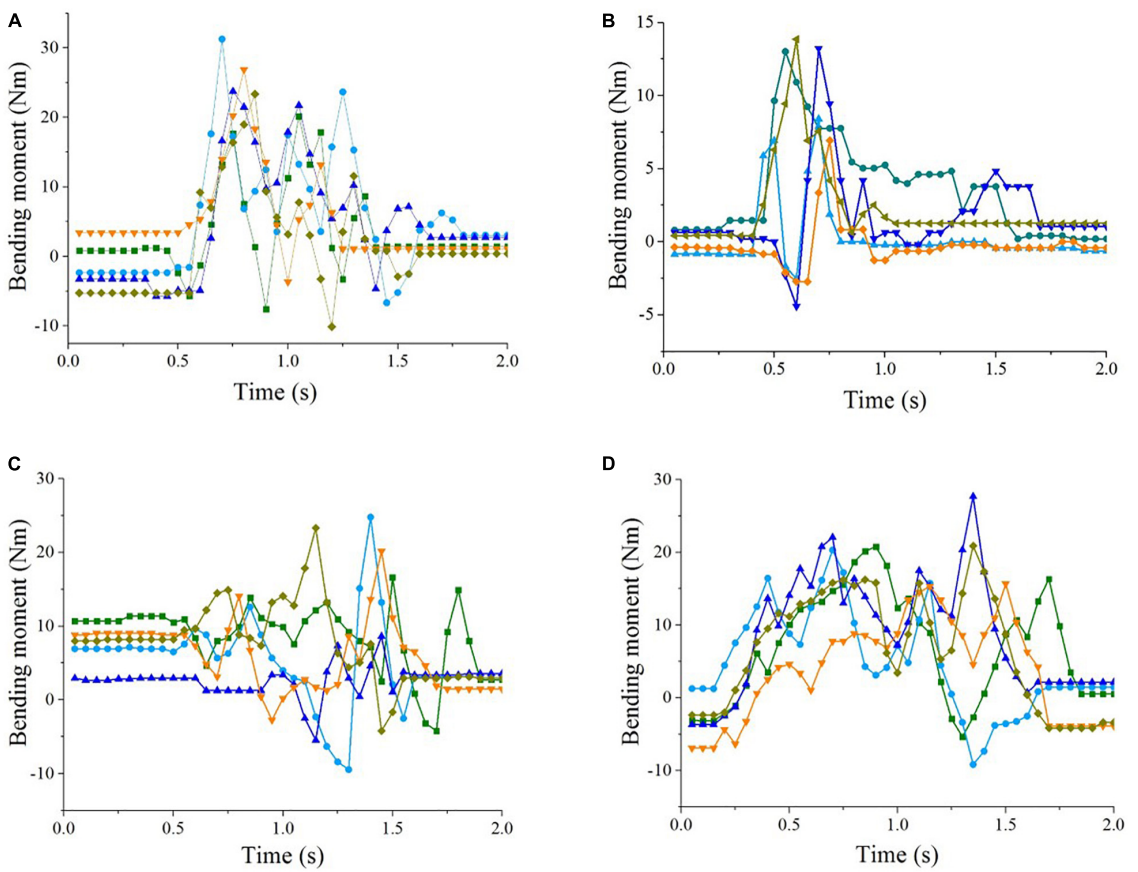


FIGURE 10 | Typical stalk bending moment curves under tested harvesting conditions. **(A)** Upright stalks; **(B)** forward lodged stalks; **(C)** reverse lodged stalks; **(D)** lateral lodged stalks.

ear was prone to be miss picked, and the header was prone to be blocked, as shown in **Figure 11D**. This result differed from the conclusions of Xue et al. (2020a) that the lowest loss occurred when the corn was harvested in the reverse direction of lodging. The above difference may be caused by the simultaneous presence of stalk breaking and root lodging in their test, which disturbed stalk feeding in the forward and lateral directions caused severe header blockage.

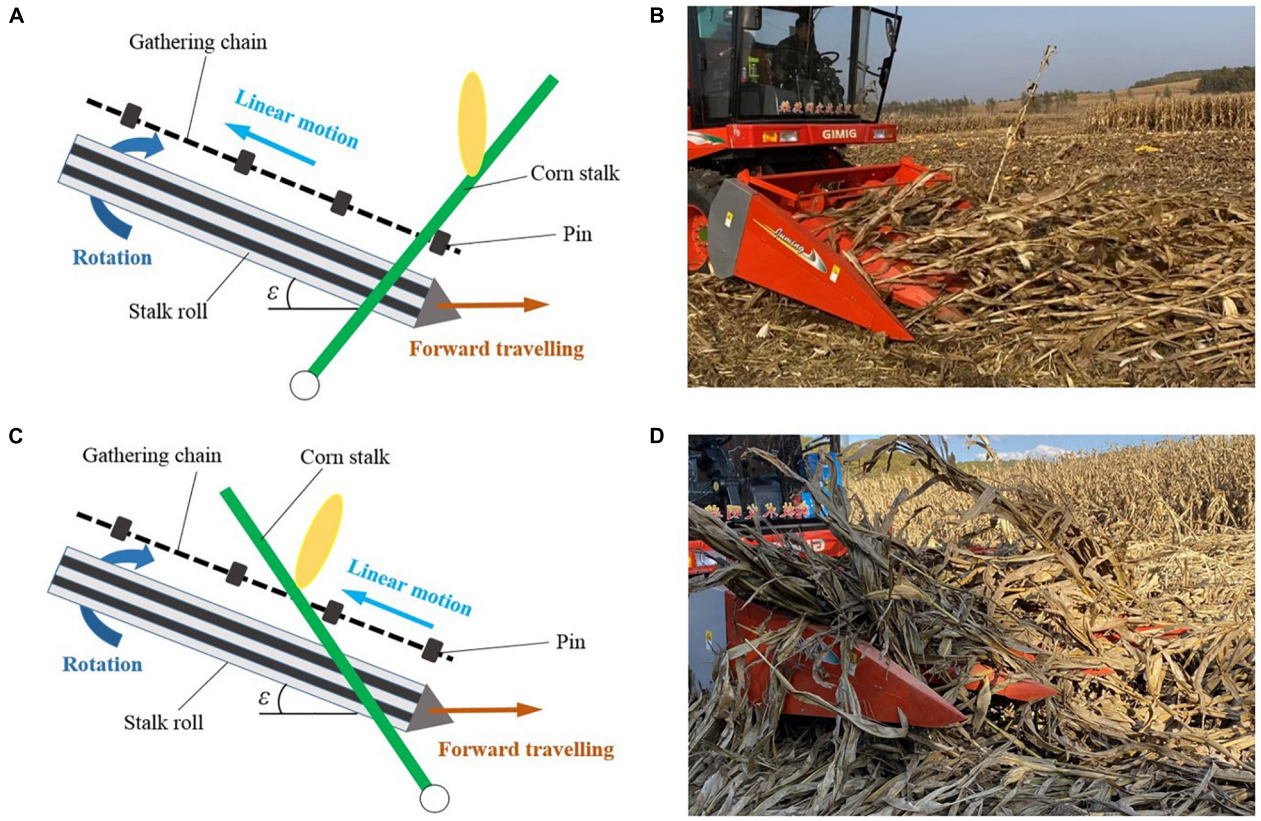


FIGURE 11 | Corn harvested in the forward and reverse directions of lodging. **(A)** Motion relationship between corn header and forward lodged corn stalks; **(B)** corn harvested smoothly in the forward direction; **(C)** motion relationship between corn header and reverse lodged corn stalks; **(D)** header blockage in the harvesting of reverse lodged corn.

In fact, in the feeding of upright corn, the velocity of the stalks could be regarded as equal to the absolute horizontal velocity of pins on the gathering chains. To avoid the stalks being pushed forward by the corn header, the absolute velocity of the pins on the gathering chains should be in the opposite direction with harvester traveling, as shown in the following formula:

$$V_{pinah} = V_{harvester} - V_{pin}\cos\varepsilon < 0 \quad (1)$$

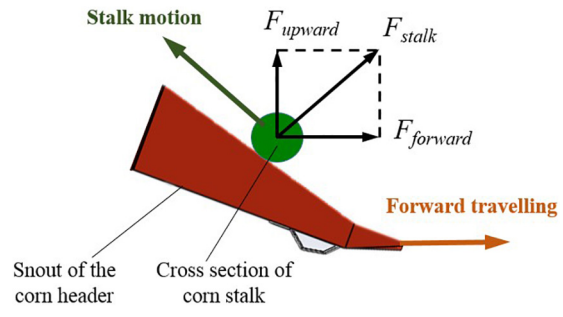


FIGURE 12 | Forces on the lateral lodged corn stalks in the lifting.

Where V_{pinah} was the absolute horizontal velocity of the pin, $V_{harvester}$ was the travel velocity of the harvester, v_{pin} was the linear motion velocity of pins driven by the gathering chains, and ε was the inclination angle of corn header.

But in the ear picking of reverse lodged corn, it needs to push the stalks forward to lift them. The forces on the stalks should be applied in the forward direction of the harvester. The absolute velocities of the pins on the gathering chains should be in the same direction with the harvester traveling:

$$V_{pinah} = V_{harvester} - V_{pin}\cos\varepsilon > 0 \quad (2)$$

The formulas (1) and (2) revealed why the corn header blockage and severe losses occurred when the harvester worked in the opposite direction with stalk lodging. In the harvest of sugarcane, the highest loss rate occurred when it was harvested opposite to the lodging direction (Tamaki et al., 2009). Because corn is harvested for the ears rather than the stalks, the corn stalks need to be pulled down by the stalk rolls after feeding. Therefore, the forces applied to the stalks on the corn header have a greater impact on corn harvesting (Zhang et al., 2021). Theoretically, to improve the corn header's ear picking adaptability for the reverse lodged corn, the velocity of pins on the gathering chains should be largely reduced to help with the lifting and feeding of stalks.

The speed-adjustable gathering chains on the corn header would be a feasible method for reverse lodged corn harvest.

In the harvesting of the lateral lodged corn stalks, a forward component force was applied to the stalk to slip it horizontally. Stalk slipping in front of the corn header may cause stalk stacking and header blockage. The component forces on a stalk are shown in **Figure 12**. The forces satisfied the following relationship:

$$F_{forward} = F_{upward} \tan \theta \quad (3)$$

Where $F_{forward}$ was the component force in the harvester forward direction, F_{upward} was the component force to lift the stalk, θ was the inclination angle of the snout surface, which was smaller than 30° (Chinese Academy of Agricultural Mechanization Sciences, 2007). It could be calculated that the $F_{forward}$ was smaller than half of the F_{upward}, and smaller than the stalk's slipping force. The stalk would not slide on the ground as long as it is not stuck on the snout surface.

Therefore, to reduce header blockage caused by stalk stacking in lateral lodged corn harvesting, the following requirements should be satisfied in the header improved design. The inclination angle of the snouts should be less than the friction angle between snouts and the stalks; the tips of the snouts should be tapering and long enough to help it enter the stalk's bottom; irregular shapes should be removed from the snout surface to avoid stalk stacking.

CONCLUSION

In corn harvesting, the torsional moments applied to the upright and lodged corn stalks in the forward, reverse, and lateral directions by the corn harvester header were 1.32, 1.59, 1.55, and 1.77 Nm, respectively, much smaller than the torsional failure boundaries 4.02 and 3.33 Nm of the upright stalks and lodged corn stalks, respectively. The bending moments applied to the upright stalks and lodged corn stalks in the forward, reverse, and lateral directions were 10.68, 22.24, 16.56, and 20.42 Nm, respectively, while the moments to push down the upright stalks, to lift the lodged stalks, and to slip the lodged stalks were 35.12, 23.33, and 40.36 Nm, respectively. The bending force was the main factor that broke the root anchorage on the corn stalks and influenced stalk lifting.

Enough bending moment applied to the stalk was an important prerequisite for lodged corn stalks lifting and corn ear picking. To improve the ability of the corn header in lifting the lodged corn stalks by applying bending moments, suggestions for improvement were made corresponding to the stalks lodged in the forward, reverse, and lateral directions. In the harvesting of forward lodged corn, the size of pins on the gathering chains needed to be enlarged to improve the corn header feeding efficiency. For the reverse lodged corn, to avoid gathering chains applying forces opposite to the stalk lifting direction, it was necessary to slow down the speed of the gathering chains to accelerate stalk lifting. In the harvesting of lateral lodged corn, lengthening and tapering the snouts and sleeking the snouts surface could help them enter the stalk bottom smoothly and avoid header blockage.

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

QF: conceptualization, methodology, data curation, and writing—original draft preparation. JF: conceptualization, resources, writing—review and editing, funding acquisition, and formal analysis. ZC: resources, methodology, software, and project administration. CC and JZ: validation, data curation, writing—review and editing, and project administration. LR: formal analysis and proofreading. All authors have read and agreed to the published version of the manuscript.

FUNDING

The research presented in this article was supported by the National Natural Science Foundation of China (Grant No. 52105257) and the Major State Research Development Program of China (Grant No. 2017YFD0700302).

ACKNOWLEDGMENTS

We would like to thank Shandong Juming Company and the Experimental Base of Agriculture of Jilin University for providing the experimental equipment.

REFERENCES

Albrecht, K. A., Martin, M. J., Russel, W. A., Wedin, W. F., and Buxton, D. R. (1986). Chemical and in vitro digestible dry matter composition of maize stalks after selection for stalk strength and stalk-rot resistance. *Crop Sci.* 26, 1051–1055. doi: 10.2135/cropsci1986.0011183X002600050043x

Bai, J., Ma, S., Wang, F., Xing, H., Ma, J., and Hu, J. (2021). Field test and evaluation on crop dividers of sugarcane chopper harvester. *Agric. Biol. Eng.* 14, 118–122. doi: 10.25165/j.ijabe.20211401.5621

Chinese Academy of Agricultural Mechanization Sciences (2007). "Grain harvester," in *Design Handbook of Agricultural Machinery*, (Beijing, China: Chinese Academy of Agricultural Mechanization Sciences).

Chauhana, S., Darvishzadeha, R., Boschettib, M., Pepe, M., and Nelsona, A. (2019). Remote sensing-based crop lodging assessment: current status and perspectives. *ISPRS J. Photogr. Remote Sens.* 151, 124–140. doi: 10.1016/j.isprsjprs.2019.03.005

Donovan, L. S., Jui, P., Kloek, M., and Nicholls, C. F. (1982). An improved method of measuring root strength in corn (Zea mays L.). *Can. J. Plant Sci.* 62, 221–223. doi: 10.3390/s18061852

Echezona, B. C. (2007). Corn-stalk lodging and borer damage as influenced by varying corn densities and planting geometry with soybean (Glycine max. L. Merrill). *Int. Agrophys.* 21, 133–143. doi: 10.1016/j.indcrop.2006.08.001

Faisal, T. R., Hristozov, N., Western, T. L., Rey, A., and Pasini, D. (2017). The twist-to-bend compliance of the Rheum rhabarbarum petiole: integrated

computations and experiments. *Comput. Meth. Biomech. Biomed. Eng.* 20, 343–354. doi: 10.1080/10255842.2016.1233328

Francisco, J., Castillo-Ruiz, S. T., and Farinelli, D. (2018). Olive fruit detachment force against pulling and torsional stress. *Span. J. Agric. Res.* 16:e0202. doi: 10.5424/sjar/2018161-12269

Fu, Q., Fu, J., Chen, Z., Han, L., and Ren, L. (2019). Effect of impact parameters and moisture content on kernel loss during corn snap-ping. *Int. Agrophys.* 33, 493–502. doi: 10.31545/intagr/113490

Han, L., Yang, G., Feng, H., Zhou, C., Yang, H., Xu, B., et al. (2018). Quantitative identification of maize lodging-causing feature factors using unmanned aerial vehicle images and a nomogram computation. *Remote Sens.* 10:1528. doi: 10.3390/rs10101528

Liu, X., Gu, W., Li, C., Li, J., and Wei, S. (2021). Effects of nitrogen fertilizer and chemical regulation on spring maize lodging characteristics, grain filling and yield formation under high planting density in Heilongjiang Province, China. *J. Integr. Agric.* 20, 511–526. doi: 10.1016/S2095-3119

Ma, D., Xie, R., Liu, X., Niu, X., Hou, P., Wang, K., et al. (2014). Lodging-related stalk characteristics of maize varieties in China since the 1950s. *Crop Sci.* 54, 2805–2814. doi: 10.2135/cropsci2014.04.0301

Martinez-Vazquez, P. (2016). Crop lodging induced by wind and rain. *Agric. For. Meteorol.* 22, 265–275. doi: 10.1016/j.agrformet.2016.07.003

Paulsen, M. R., Pinto, F. A. C., Sena, D. G., Zandonadi, R. S., Ruffato, S., Costa, A. G., et al. (2014). Measurement of combine losses for corn and soybeans in Brazil. *Appl. Eng. Agric.* 30, 841–855. doi: 10.13031/aea.30.10360

Phetmanyseng, X., Senthong, P., Chay, B., and Shu, F. (2019). Combine harvesting efficiency as affected by rice field size and other factors and its implication for adoption of combine contracting service. *Plant Prod. Sci.* 22, 68–76. doi: 10.1080/1343943X.2018.1561196

Reneau, J. W., Khangura, R. S., Stager, A., Erndwein, L., Weldekidan, T., Cook, D. D., et al. (2020). Maize brace roots provide stalk anchorage. *Plant Direct* 4:e00284. doi: 10.1002/pld3.284

Robertson, D. J., Lee, S. Y., Julias, M., and Cook, D. D. (2016). Maize stalk lodging: flexural stiffness predicts strength. *Crop Sci.* 56, 1711–1718. doi: 10.2135/cropsci2015.11.0665

Seegmiller, W. H., Graves, J., and Robertson, D. J. (2020). A novel rind puncture technique to measure rind thickness and diameter in plant stalks. *Plant Meth.* 16:44. doi: 10.1186/s13007-020-00587-4

Sekhon, R. S., Joyner, C. N., Ackerman, A. J., McMahan, C. S., Cook, D. D., and Robertson, D. J. (2020). Stalk bending strength is strongly associated with maize stalk lodging incidence across multiple environments. *Field Crops Res.* 249:107737. doi: 10.1016/j.fcr.2020.107737

Shelby, R. J., Villamil, M. B., Ames, K. A., and Bradley, C. A. (2018). Effects of pyraclostrobin foliar fungicide, corn hybrid, and harvest timing on stalk health of corn. *Crop Forage Turfgrass Manag.* 4, 1–5. doi: 10.2134/cftm2018.02.0006

Stubbs, C. J., Cook, D. D., and Niklas, K. J. (2019). A general review of the biomechanics of root anchorage. *J. Exp. Bot.* 70, 3439–3451. doi: 10.1093/jxb/ery451

Tamaki, M., Shikanai, T., Akachi, T., and Adaniya, A. (2009). Investigation of missing plant for cane plant and development of supplemental planter. *J. JAPAN Soc. Agric. Mach.* 71, 104–114. doi: 10.11357/jsam.71.3_104

Wang, G., Jia, H., Tang, L., Zhuang, J., Jiang, X., and Guo, M. (2016). Design of variable screw pitch rib snapping roller and residue cutter for corn harvesters. *Int. J. Agric. Biol. Eng.* 9, 27–34. doi: 10.3965/j.ijabe.20160901.1941

Wang, K. R., Xie, R. Z., Ming, B., Hou, P., Xue, J., and Li, S. K. (2021). Review of combine harvester losses for maize and influencing fac-tors. *Agric. Biol. Eng.* 14, 1–10. doi: 10.25165/j.ijabe.20211401.6034

Wilke, N., Siegmann, B., Klingbeil, L., Burkart, A., Kraska, T., Muller, O., et al. (2019). Quantifying lodging percentage and lodging severity using a UAV-based canopy height model combined with an objective threshold approach. *Remote Sens.* 11:515. doi: 10.3390/rs11050515

Windsor-Collins, A. G., Atherton, M. A., Collins, M. W., and Cutler, D. F. (2007). Section properties of palm petioles. Part 1: the influence of shape on the flexural and torsional properties of selected palm petioles. *Int. J. Des. Nat.* 2, 328–347.

Xue, J., Dong, P., Hu, S., Li, L., Wang, K., Gao, S., et al. (2020a). Effect of lodging on maize grain loss and loss reduction technology in mechanical grain harvest. *J. Maize Sci.* 28, 116–120. doi: 10.13597/j.cnki.maize.science.20200617

Xue, J., Ming, B., Xie, R., Wang, K., Hou, P., and Li, S. (2020b). Evaluation of maize lodging resistance based on the critical wind speed of stalk breaking during the late growth stage. *Plant Meth.* 16:148. doi: 10.1186/s13007-020-00689-z

Yang, L., Cui, T., Qu, Z., Li, K., Yin, X., Han, D., et al. (2016). Development and application of mechanized maize harvesters. *Int. J. Agric. Biol. Eng.* 9, 15–28. doi: 10.3965/j.ijabe.20160903.2380

Zhang, Z., Chi, R., Du, Y., Pan, X., Dong, N., and Xie, B. (2021). Experiments and modeling of mechanism analysis of maize picking loss. *Int. J. Agric. Biol. Eng.* 14, 11–19. doi: 10.25165/j.ijabe.20211401.5745

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fpls.2022.849260

A Loosely Coupled Extended Kalman Filter Algorithm for Agricultural Scene-Based Multi-Sensor Fusion

*Meibo Lv, Hairui Wei, Xinyu Fu, Wuwei Wang and Daming Zhou**

School of Astronautics NPU, Northwestern Polytechnical University, Xi'an, China

With the arrival of aging society and the development of modern agriculture, the use of agricultural robots for large-scale agricultural production activities will become a major trend in the future. Therefore, it is necessary to develop suitable robots and autonomous navigation technology for agricultural production. However, there is still a problem of external noise and other factors causing the failure of the navigation system. To solve this problem, we propose an agricultural scene-based multi-sensor fusion method *via* a loosely coupled extended Kalman filter algorithm to reduce interference from external environment. Specifically, the proposed method fuses inertial measurement unit (IMU), robot odometer (ODOM), global navigation and positioning system (GPS), and visual inertial odometry (VIO), and uses visualization tools to simulate and analyze the robot trajectory and error. In experiments, we verify the high accuracy and the robustness of the proposed algorithm when sensors fail. The experimental results show that the proposed algorithm has better accuracy and robustness on the agricultural dataset than other algorithms.

Keywords: loosely coupling, extended Kalman filter algorithm, multi-sensor fusion, robustness, agricultural robot

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Agnieszka Szczęsna,
Silesian University of Technology, Poland
Jingjie Xie,
University of Warwick, United Kingdom
Jianfeng Cui,
North University of China, China

***Correspondence:**
Daming Zhou
daming.zhou@nwpu.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *05 January 2022*
Accepted: *14 February 2022*
Published: *25 April 2022*

Citation:
Lv M, Wei H, Fu X, Wang W and Zhou D (2022) A Loosely Coupled Extended Kalman Filter Algorithm for Agricultural Scene-Based Multi-Sensor Fusion. Front. Plant Sci. 13:849260. doi: 10.3389/fpls.2022.849260

INTRODUCTION

In recent years, with the development of artificial intelligence technology, agricultural robots such as drones and ground mobile carts (Katsigiannis et al., 2016; Tang et al., 2020; Atefi et al., 2021; Qin et al., 2021) have been gradually applied to modern agriculture. Their ability to sense the environment and navigate on their own is a more critical influencing factor. And multi-sensor fusion technology provides an effective method for agricultural robots to enhance their ability to work in complex and uncertain environments (Noguchi et al., 1998; Viacheslav et al., 2011).

Multi-sensor fusion technology is a multi-level complementary. It optimally processes the information from different types of sensors to form a reasonable interpretation of the observed environment. Compared with the traditional single-sensor technology, it is fault-tolerant, complementary, real-time, economical, and can solve the defects caused by single sensor, such as fuzzy points and so on. And all of these allow it a more accurate observation of the environment. Therefore, multi-sensor fusion technology has received wide attention in various fields such as military, control, and signal processing (Abidi and González, 1992; Hall and Llinas, 1997; Varshney, 2000).

Visual inertial odometry (VIO) is an application of multi-sensor fusion technology. At present, the mainstream VIOs includes VINS_MONO (Tong et al., 2018), VINS_FUSION (Tong et al., 2018) and MSCKF_VIO (Mourikis and Roumeliotis, 2007). They are used to accomplish map construction, navigation and positioning functions by fusing visual sensors and inertial measurement units (IMUs). According to the difference of fusion framework,

vision inertial odometry can be further divided into two types: tightly coupled and loosely coupled. In the loosely coupled (Faessler et al., 2015, 2016; Delmerico and Scaramuzza, 2018), the visual motion and inertial navigation system has two independent modules. In addition, the filter is used to decouple and fuse the visual and IMU information, which has the characteristic of simplicity and speed. VINS-MONO is an open source VIO algorithm, which is realized by tight coupling method and restores the scale through Monocular and IMU. VINS-FUSION is an optimization-based multi-sensor state estimator, it achieves accurate self-localization for autonomous applications (drones, cars, and AR/VR). VINS-FUSION is an extension of VINS-MONO, which supports multiple visual-inertial sensor types (MONO camera + IMU, stereo cameras + IMU, even stereo cameras only). MSCKF_VIO is a binocular visual odometry based on multi-state constraint Kalman filter. Multi-state constraint refers to adding the camera pose of multi-frame images to the Kalman state vector by using least square optimization and estimating 53 the spatial position of feature points through the constraints between multi frame images. Then the state vector is constrained based on the spatial position of the optimized feature points. In the field of agricultural robots, the research on machine vision and trajectory navigation is gradually deepening, and visual-inertial navigation combined with other methods is constantly evolving.

The core of VIO algorithm is based on the state optimization of filtering methods (Scaramuzza et al., 2014), where the filtering methods are mainly based on Bayesian estimation theory, including Kalman filter (KF) algorithm (Gao et al., 2017) and particle filtering algorithm (Leutenegger et al., 2015), etc. Among these, the KF algorithm is used more widely in practical applications. Since SLAM methods are generally non-linear when performing system observations and measurements, KF algorithms need to be extended to the non-linear domain. Accordingly, researchers propose the Extended Kalman Filter (EKF) (Kalman, 1960), a linear approximation method in ignoring higher-order terms for non-linearity, which can estimate the state of a dynamic system from a series of measurements that do not exactly contain noise. Though it is a suboptimal filtering algorithm, it solves the problem of nonlinear systems in the KF algorithm. EKF is widely used in the field of robotics. An improved covariance Intersection EKF data fusion algorithm is proposed for multi-sensor time-delay system (Lee and Gao, 2019). A slam method with extended Kalman filter (EKF) is introduced to locate landmark robots and mobile robots (Inam et al., 2020).

Many studies on multi-sensor fusion algorithms for agricultural robots have been conducted by reasearchers. At the end of last century, many scholars proposed vision based automatic agricultural machine perception, navigation obstacle avoidance and other related methods (Ollis and Stentz, 1997; Sharma and Borse, 2016; Reina and Messina, 2019). Recently, research has developed rapidly, including developed modular structured robots that use GPS for navigation and positioning, and multi-sensor fusion for robot obstacle avoidance (Liu et al., 2011). A multi-sensor data fusion algorithm has been presented based on the fusion set, it is mainly used for data collection in agricultural systems (Hu and Yan, 2018). A multi-sensor fusion approach has been developed for autonomous navigation of agricultural vehicles, which is applied for crop row tracking and traversable operations (Benet and Lenain, 2017). Along with the great improvement in integrated navigation and sensor fusion, a class of autonomous driving control algorithms has been proposed in order to achieve high-precision autonomous navigation of tracked agricultural vehicles, which includes GNSS-RTK sensor integration algorithm and path tracking algorithm (Han et al., 2020). A LiDAR-based autonomous navigation system is developed for agricultural robots, which fuses LiDAR and IMU to solve the problem of agricultural navigation when the tree canopy is obscured (Velasquez et al., 2021). At present, the research of agricultural robots combined with multi-sensor fusion technology is in a rapid development stage. However, the sensor technology generally relies too heavily on GNSS or GPS navigation system, and the sensor fusion category is single, generally using only two sensors for fusion, which may lead to a failure of the whole system when the navigation system has problems.

This paper presents a multi-sensor fusion algorithm based on a loosely coupled extended Kalman filter, the proposed method reincorporates the robot odometer (ODOM), global navigation and positioning system (GPS), and the inertial measurement unit (IMU) on the top of the visual odometer for agricultural robots. And due to the favorable features of GPS navigation such as wide coverage, strong resistance to climate influence and real-time dynamics (RTK), we introduce a loosely coupled and extended Kalman filtering algorithm to fuse the GPS and VINS-MSCKF, ODOM, and IMU. In addition, the effects of GPS or sensors failure on the system are also analyzed. Based on the analysis, it is obvious that the proposed algorithm can better solve the problem for the system downtime situation caused by the failures of GPS and VIO sensors. Based on these experimental results, we can conclude that the proposed algorithm can effectively improve navigation accuracy and system robustness under agricultural scenes.

EXTENDED KALMAN FILTERING AND MULTI-SENSOR FUSION REVIEW

Extended Kalman Filter

As a linearized approximation method, extended Kalman filtering (Sastry, 1971) is a class of extended form of standard Kalman filtering in nonlinear systems.

$$\begin{cases} x_{k+1} = f(x_k) + w_k \\ z_k = h(x_k) + v_k \end{cases} \tag{1.1}$$

where x_k and z_k are the state vector and the measurement vector, w_k and v_k are system noise and measurement noise, respectively, with covariance Q_k,R_k. The system state equation is:

$$x_{k+1} = f\left(\hat{x}_{k\vee k}\right) + F_k\left(x_k - \hat{x}_{k\vee k}\right) + w_k \tag{1.2}$$

where, $F_k = \partial f / \partial x_k|_{x_k=\hat{x}_{k|k}}$

One-step state prediction equation:

$$\hat{x}_{k+1\vee k} = E\left[x_{k+1} + w_k\right] = f\left(\hat{x}_{k+|k}\right) \tag{1.3}$$

The one-step state prediction covariance is:

$$P_{k+1\vee k} = F_k P_{k|k} F_k^T + Q_k \tag{1.4}$$

One-step measurement prediction equation:

$$\hat{z}_{k+1|k} = E\left[h_{k+1} + v_k\right] = h\left(\hat{x}_{k+1|k}\right) \tag{1.5}$$

Measurement prediction error covariance array:

$$P_{zz,k+1|k} = H_{k+1} P_{k+1|k} H_{k+1}^T + R_{k+1} \tag{1.6}$$

The reciprocal covariance matrix between the state and the measurement equations:

$$P_{xz,k+1|k} = P_{k+1|k} H_{k+1}^T \tag{1.7}$$

State gain matrix:

$$K_{k+1} = P_{k+1} H_{k+1}^T \left(H_{k+1} P_{k1|k} H_{k+1}^T + R_{k+1}\right)^{-1} \tag{1.8}$$

The state estimates at moment k+1 is:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|1} + K_{k+1}(z_{k+1} - \hat{z}_{k+1|k} \tag{1.9}$$

State estimation error covariance matrix is:

$$P_{k+1\vee k+1} = \left(I - K_{k+1} H_{k+1}\right) P_{k+1|k} \left(I - K_{k+1} H_{k+1}\right)^T + K_{k+1} R_{k+1} {K_{k+1}}^T \tag{1.10}$$

Eqs. (1.2) to (1.10) form the extended Kalman filter algorithm.

In this paper, we define the failure state of GPS and VIO. When there is no GPS signal, it is defined as GPS failure. When the distance between the two adjacent VIO frames is greater than a given threshold, it is defined as the VIO failure. During the GPS or VIO fails, we perform sensor fusion by discarding the GPS or VIO state variable values, and replace the failed GPS or VIO values by the wheel odometer's position, quaternion, and covariance values. Compared with the traditional Kalman filter algorithm, the proposed loosely coupled extended Kalman filter algorithm can perform tasks when the GPS or VIO fails, since the fusion mechanism includes the failure judgment of GPS and VIO. Therefore, the judgment of the fusion mechanism can remove the influence of the failed GPS or the invalid VIO on the whole system, and thus replace the state variables of a failed GPS or a failed VIO with the wheel odometer's position, attitude, quaternion, and covariance values. Therefore, the system can still operate normally even when GPS or VIO fails.

The mathematical description of this paper is based on the extended Kalman filter theory, and the state variables involved are the position and attitude of GPS, wheel odometer, VIO, and the attitude of IMU. From a mathematical point of view, the proposed method is to switch the state variable to the value of the wheel odometer by judging the GPS signal state and the position change of VIO, in order to achieve system stability.

Multi-Sensor Fusion

Multi-sensor fusion (MSF) currently completes the required measurement estimates for subsequent information processing by using computer technology. In this way, automatically analyze and synthesize data from multiple sensors or multiple sources with certain criteria.

The multi-sensor fusion is characterized by complexity and multi-level, and its basic requirements for algorithms are robustness, parallel processing capability, speed and accuracy of operations, etc. It is also necessary to consider the performance of the connection with the previous pre-processing (information input) and subsequent information processing system (system output), etc. In general, mathematical methods based on non-linearity and with features such as fault tolerance and adaptability can be used as fusion algorithms. At present, most of the research on sensor fusion algorithms based on Kalman filter include adaptive Kalman filter, extended Kalman filter, volumetric Kalman filter and unscented Kalman filter. In these studies, the model parameters and the system noise characteristics can be estimated and updated only when the sensor is working normally. When the sensor fails, the whole system collapse. In this paper, the failure judgment is added on the basis of the past, and a loosely coupled EKF algorithm is proposed to make the system run normally.

MATERIALS AND METHODS

Algorithm Framework

The proposed algorithm framework is shown in **Figure 1**. The input of algorithm are images captured by the binocular camera and the measurement from the IMU. And the information about

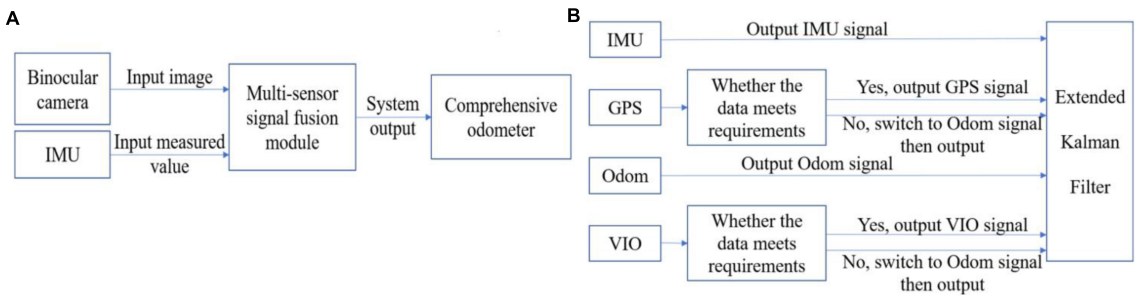


FIGURE 1 | (A) Overall algorithm framework and **(B)** multi-sensor signal fusion module.

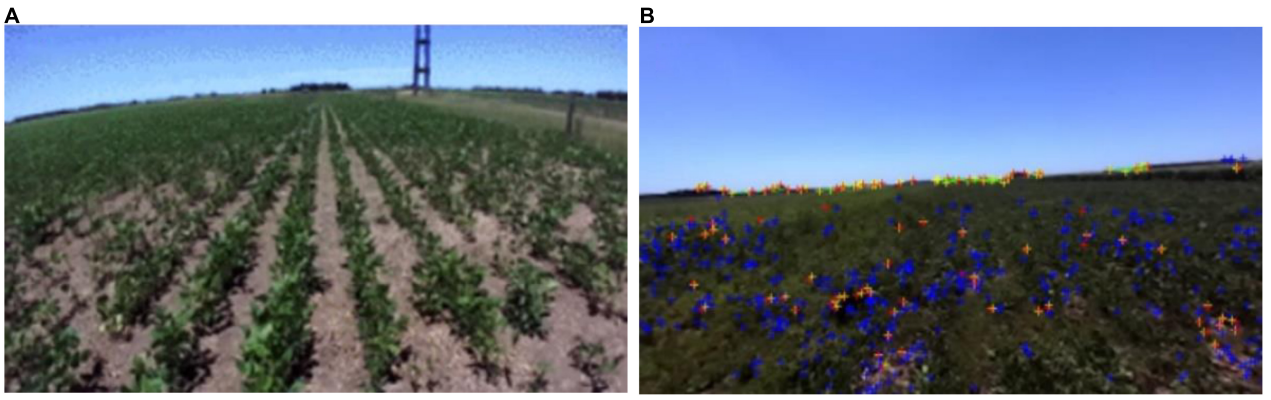


FIGURE 2 | (A) Data set–repeated scenes and **(B)** example of dataset tracking.

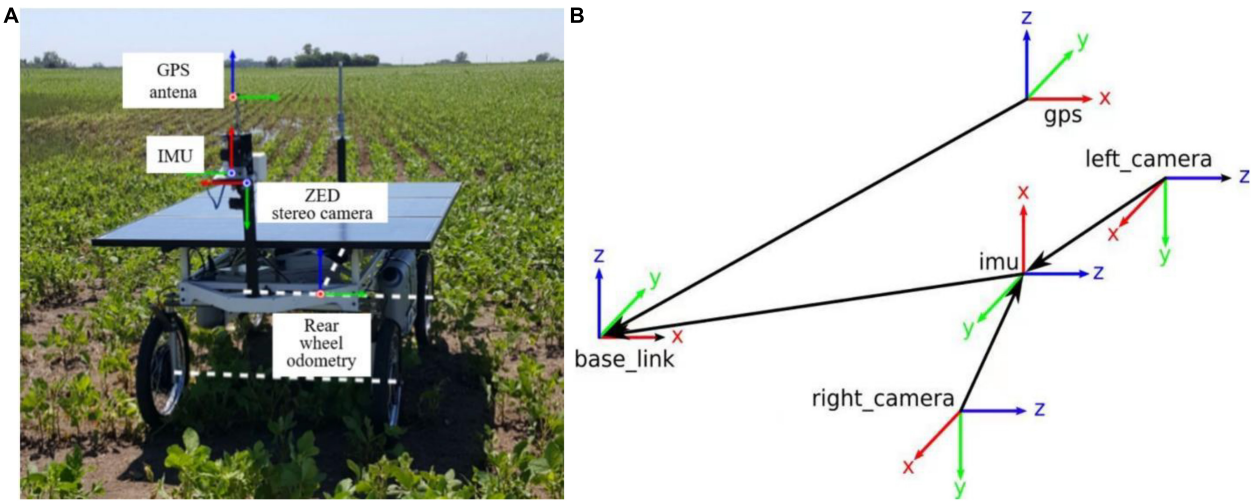


FIGURE 3 | (A) Weeding mobile robot equipped with sensors and **(B)** dataset coordinate system definition.

the farming system, i.e, the farming data set, is obtained from the input binocular image. At the same time, the position pose estimation of the farming robot is completed based on the proposed fusion algorithm.

The extended Kalman filter fusion algorithm cannot obtain stable data since GPS and VIO are greatly affected by the environment. For example, GPS basically cannot receive satellite signals after being blocked, and VIO is too sensitive to ambient light. To solve this problem, the proposed algorithm fuses four sensors, among which the data from GPS and VIO need to be evaluated. The algorithm can perform decision-level fusion by adding a condition to determine whether the sensor fails. The algorithm judges whether the GPS is invalid through the differential state. When the output of the differential positioning state is 2, the GPS works normally. Otherwise, the algorithm will replace the GPS data with the ODOM data. The same goes for VIO. The algorithm determines whether the distance between the two adjacent VIO frames is greater than a given threshold. If the VIO fails, the proposed algorithm will use the ODOM data instead of the VIO data. When the sensor returns to normal, the data of GPS and VIO are re-added to the fusion system for fusion to correct the system error. After synchronizing the time of each sensor, the algorithm uses the EKF filtering formula to process the data, and outputs the attitude estimation value of the whole robot.

The data fused in the algorithm contain the odometer information converted by GPS through coordinates, including the covariance and coordinate values for x and y axes of the GPS; the covariance and coordinate values for x and y axes of the ODOM; the quaternion and covariance of the IMU; the covariance and coordinate values for x, y, and z axes of the VIO and the quaternion output by the tight coupling between the IMU and the camera. And all these data are obtained in the carrier coordinate system ($O_bX_bY_bZ_b$). In addition, the coordinate systems considered in this paper include: Geographic Coordinate System ($O_gX_gY_gZ_g$), Camera Coordinate System ($O_cX_cY_cZ_c$), Navigation Coordinate System ($O_nX_nY_nZ_n$), Inertial Coordinate System ($O_iX_iY_iZ_i$) and Pixel Coordinate System (O_puv).

The proposed fusion of sensor information framework is shown in **Figure 1**, it can improve the navigation accuracy while enhancing the robustness of the system effectively. The

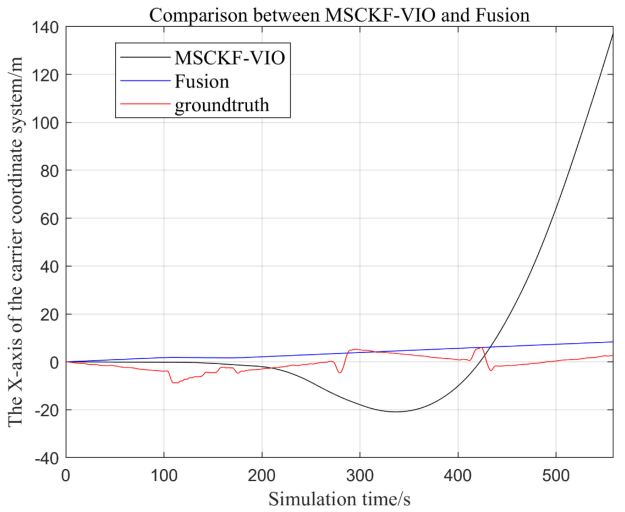


FIGURE 4 | Comparison of the overall trajectory of the fusion algorithm with the MSCKF_VIO algorithm in the *x*-axis.

data analysis and final results are given in the experiments and results section.

Signal Fusion

In the signal fusion stage, we fuse the signals from IMU, ODOM, VIO, and GPS inputs based on the extended Kalman filtering algorithm in a loosely coupled manner. Then the fusion algorithm is used to estimate the real position and attitude information of the ground farm robot, and outputs the fused and filtered combined odometer (ODOM combined) information. The update of the positional attitude information, covariance information, and timing update are introduced as follows.

(1) Position-attitude update.
All sensor sources have their own reference coordinate system and may drift with time. To solve this problem, we replace the absolute position pose information with the relative position pose difference.
(2) Covariance update.
As the robot moves over a larger and larger range, the uncertainty of its position pose gradually increases, the covariance increases accordingly, and the absolute covariance of position pose become less meaningful; therefore, the sensors release a period of covariance change to update the covariance, i.e., the covariance of velocity.
(3) Timing update.
It is assumed that the initial update moment of the farm robot to the filter is t_0. In this case, the filter subscribes to the fused position information at t_1, IMU information at t_2, ODOM information at t_3, GPS information at t_4, etc. Then the IMU, ODOM, and GPS information are interpolated at t_0 and t_1, t_0 and t_2, and t_0 and t_3, respectively. The EKF filter will use the information obtained from these linear interpolations to calculate the integrated odometer data updated by the filter at time t_1.

Algorithm Input

The Rosario Dataset (Pire et al., 2019) is a class of publicly available datasets in agriculture (**Figure 2**), this dataset is used for mobile robots in agricultural scenarios in terms of agricultural sensor fusion, SLAM, etc. This dataset is provided by a weeding mobile robot equipped with a stereo camera, GPS-RTK sensor, and IMU (**Figure 3A**) for agricultural field. **Figure 3B** represents the dataset coordinate system, which means that the world coordinate system completes the pose transformation and turns to the map coordinate system. It is assumed that the initial update moment of the farm robot to the filter is t_0. In this case, the filter subscribes to the fused position information at t_1, IMU information at t_2, ODOM information at t_3, GPS information at t_4, etc. The collected information is fused with odometry, inertial, and visual information for further processing. Consequently, the results are generated by deriving environmental data covering highly repetitive scenes, reflection and burn images, direct sunlight scenes, and rough terrain scenes.

This dataset has a relatively universal character, including a range of different farmland scenes, and is suitable for the study of this paper. For this reason, the dataset is used as the initial input dataset for the proposed algorithm. In the dataset, Ground Truth is the real motion trajectory of the robot, i.e., the standard trajectory.

EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, the Rosario Dataset is used to simulate the trajectory of the robot in the spatial Cartesian coordinate system by using the starting position of the robot as the zero point. All the data mentioned in the paper are obtained by running the proposed algorithm for the farming scenario.

Based on the unified dataset, the proposed approach analysis can be divided into the following three aspects.

(1) Judging the algorithm accuracy by comparing the trajectories of the fusion algorithm proposed in this paper with other algorithms.
(2) Judging the robustness of the proposed fusion algorithm based on the output trajectory by introducing Gaussian distribution noise to disable specific sensors.
(3) Changing the fusion judgment condition of VIO algorithm, output trajectory and observe the influence of different judgment conditions on the proposed fusion algorithm.

It should be noted that in the following simulation diagrams, the X direction represents the left and right transverse direction of the vehicle body, Y represents the front and rear longitudinal direction of the vehicle body.

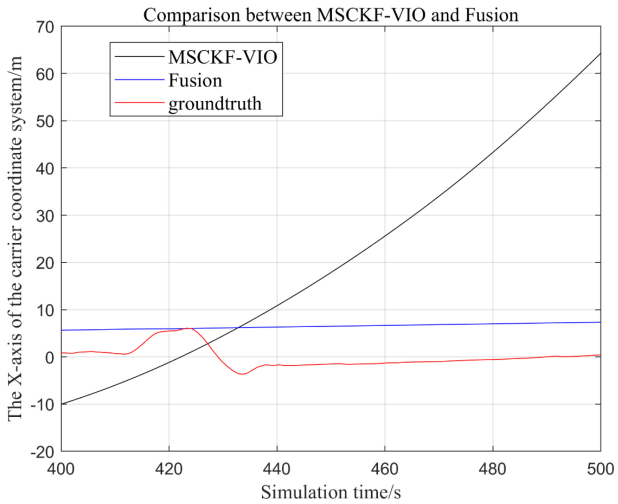


FIGURE 5 | Comparison of the fusion algorithm with the MSCKF_VIO algorithm for zooming in on the *x*-axis local trajectory.

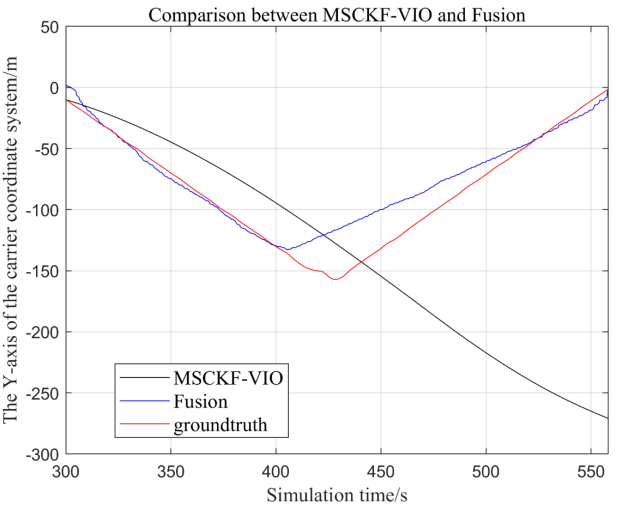


FIGURE 7 | Zoomed-in comparison of the *y*-axis local trajectories of the fusion algorithm and the MSCKF_VIO algorithm.

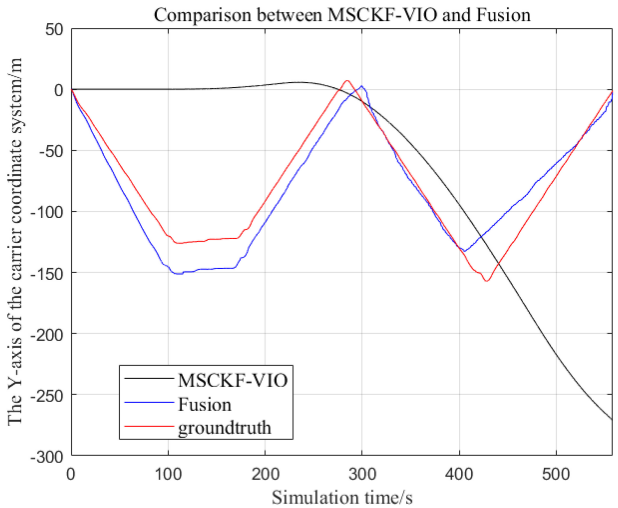


FIGURE 6 | Comparison of the overall trajectory in *y*-axis between the fusion algorithm and the MSCKF_VIO algorithm.

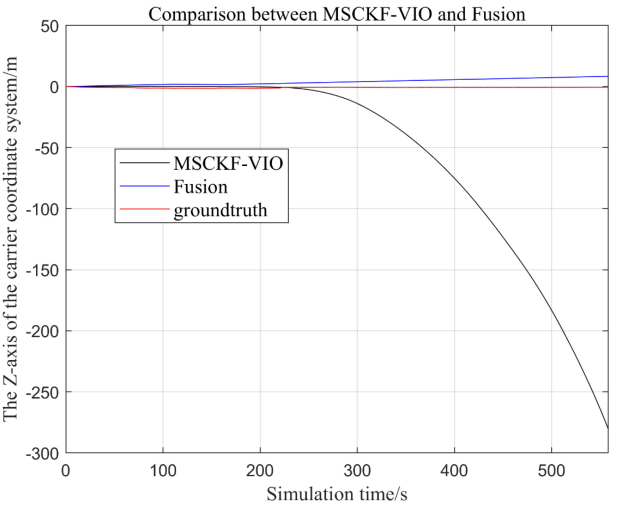


FIGURE 8 | Comparison of the overall trajectory in *z*-axis between the fusion algorithm and MSCKF_VIO algorithm.

Experimental Simulation and Analysis for Trajectory Accuracy

Comparison of the Proposed Loosely Coupled Extended Kalman Filter Multi-Sensor Fusion Algorithm With the MSCKF_VIO Algorithm

There are three trajectories in the following figures. From **Figures 4–9**, the trajectory of loosely coupled extended Kalman filter algorithm is represented by Fusion, the trajectory of MSCKF_VIO is represented by MSCKF_VIO, and the standard trajectory is represented by ground truth, respectively. Where the **Figures 5**, **7**, **9** intercept the part of *t*-axis coordinates from 0 to 150 s in **Figures 4**, **6**, **8**.

(1) Comparison of trajectories in the *x*-axis direction.
(2) Comparison of trajectories in the *y*-axis direction.
(3) Comparison of trajectories in the *z*-axis direction.
(4) Analysis of the figures.

By analyzing **Figures 4–9**, it can be found that when the robot first starts to run, the difference of the three trajectories is not large and the error is within the acceptable range. However, with the increase of running time, the trajectory error between the MSCKF_VIO algorithm and the standard trajectory is increasing. Furthermore, the speed of error increase is also improving. In contrast, the fusion algorithm proposed in this paper reflects its superior stability and higher accuracy in the overall trajectory.

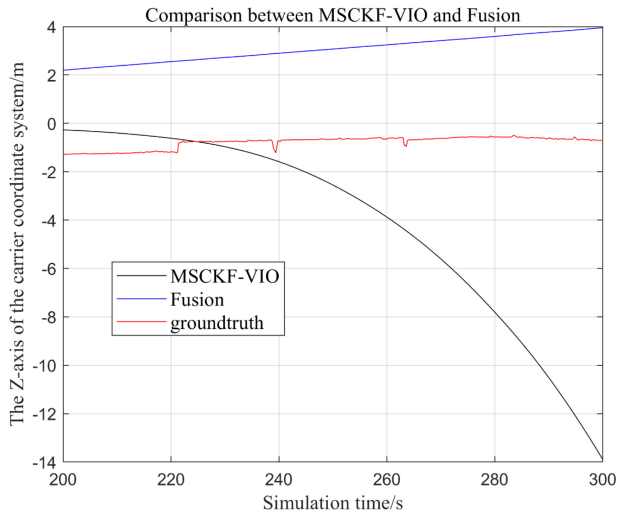


FIGURE 9 | Zoomed-in comparison of the *z*-axis local trajectories of the fusion algorithm and the MSCKF_VIO algorithm.

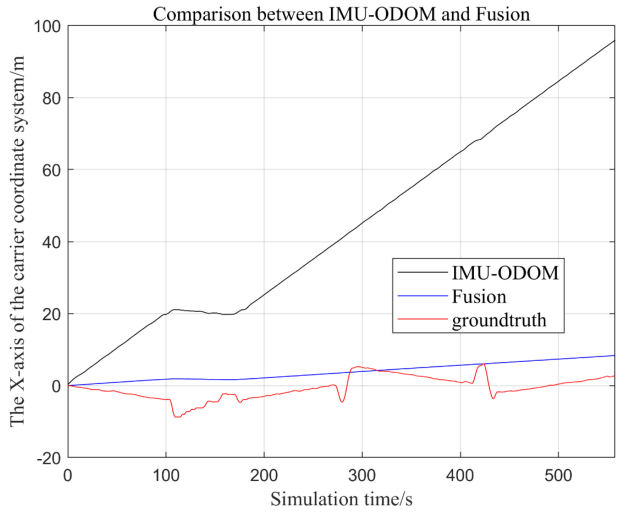


FIGURE 10 | Overall trajectory comparison of fusion algorithm with IMU and ODOM fusion algorithm in *x*-axis.

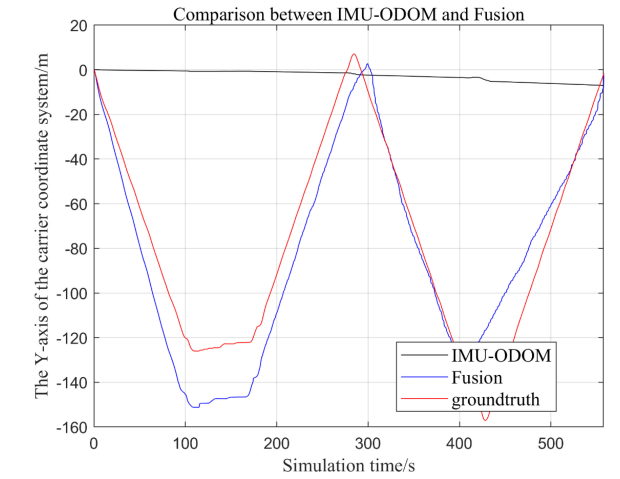


FIGURE 11 | Overall trajectory comparison of fusion algorithm with IMU and ODOM fusion algorithm in *y*-axis.

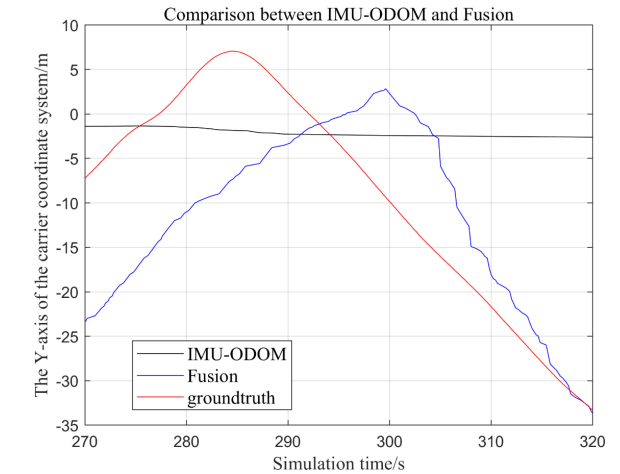


FIGURE 12 | Comparison of *y*-axis local trajectory of fusion algorithm with IMU and ODOM fusion algorithm.

Comparison of the Proposed Loosely Coupled Extended Kalman Filter Based Multi-Sensor Fusion Algorithm With Inertial Measurement Unit and ODOM Fusion Algorithm

There are three trajectories in the following figures. **Figures 10–16** show the trajectory of the proposed loosely coupled EKF algorithm (denoted as Fusion), IMU-ODOM, and the standard trajectory (denoted as ground truth), respectively by IMU-ODOM. By comparing the trajectories of the Fusion algorithm, MSCKF_VIO algorithm and IMU and ODOM fusion algorithm proposed in this paper with the standard trajectories, the accuracy of the two algorithms can be judged.

(1) To make a more accurate analysis of the results, **Figure 12** intercept the part of *t*-axis coordinates from 200 to 300 s in **Figures 11**, **14** remove the standard trajectory and intercept the part of *t*-axis coordinates from 400 to 500 s.
(2) Comparison of the trajectory in the *y*-axis direction.
(3) Comparison of trajectories in *z*-axis direction.
(4) Analysis of the figures.

By analyzing from **Figures 10–13**, in the *x*-axis trajectory, the accuracy of fusion algorithm of IMU and ODOM is obviously lower than the accuracy of the proposed fusion algorithm; in the *y*-axis trajectory, the difference in accuracy between the two is not significant. Since the car body moves in the plane, this paper only considers the *x* and *y* axes of the two-dimensional plane in the IMU-ODOM fusion simulation, and its value is zero compared to the real value in the z-axis trajectory.

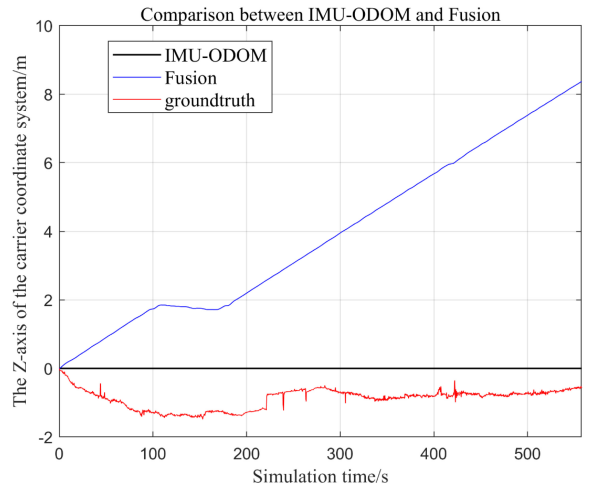


FIGURE 13 | Overall trajectory in z-axis of fusion algorithm with IMU and ODOM fusion algorithm.

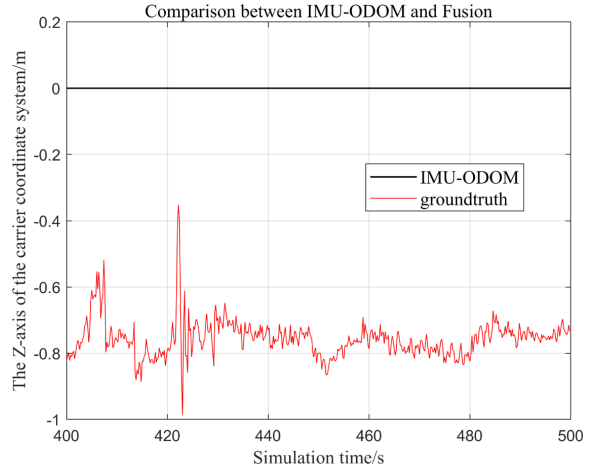


FIGURE 14 | Comparison of *z*-axis local trajectory of fusion algorithm with IMU and ODOM fusion algorithm.

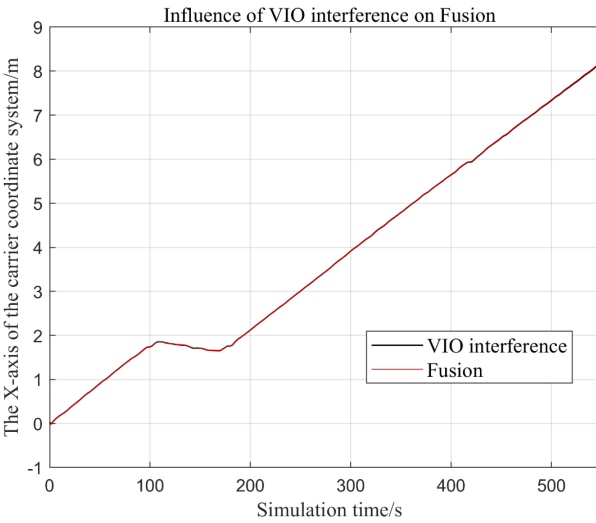


FIGURE 15 | Comparison of the overall trajectory in the *x*-axis after VIO failure and when all sensors are working normally.

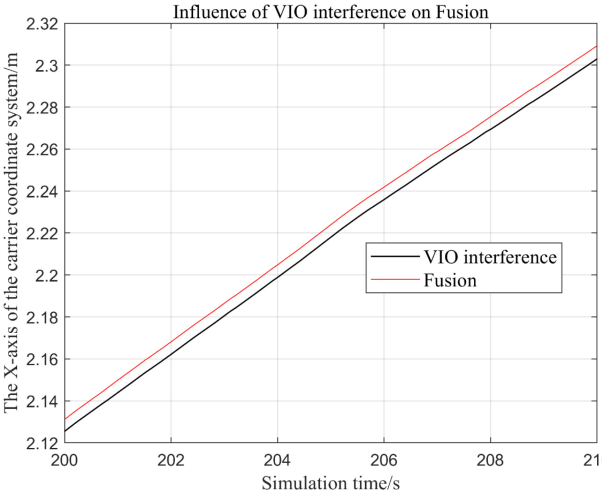


FIGURE 16 | Comparison of local trajectory in *x*-axis after VIO failure and when all sensors are working normally.

Analysis of the Tables

In this paper, the results of different sensor fusion methods are quantitatively compared with standard trajectories, the average value and mean square deviation of the absolute value are shown in **Tables 1**, **2**.

According to the conclusions drawn from the previous analysis with **Tables 1**, **2**, it is concluded that the stability and robustness of proposed fusion algorithm are significantly superior to the MSCKF-VIO algorithm and the IMU-ODOM fusion algorithm, and it can be found that the trajectory of fusion is closer to the real trajectory than those obtained by other algorithms, and thus the accuracy of the proposed fusion algorithm is better than IMU and ODOM fusion algorithm.

Experiments for the Robustness of the System

The robustness of the system reflects its characteristic of maintaining certain performance under certain parameter uptake. In this paper, the robustness of the system is verified by adding Gaussian distribution noise to disable a specific sensor, and then the trajectory with sensor disablement is compared to the trajectory without the disturbance.

In this part, we disable the sensors GPS and VIO respectively when $t \in [200, 300]$.

TABLE 1 | Average of absolute value of errors in x, y, and z directions.

Axis	Comparison		
	Fusion-groundtruth	MSCKF-groundtruth	IMU-groundtruth
x/m	4.5948	20.9995	45.2532
y/m	14.9209	75.3966	75.0033
z/m	4.7763	54.2210	0.8808

TABLE 2 | Mean square deviation of errors in x, y, and z directions.

Axis	Comparison		
	Fusion-groundtruth	MSCKF-groundtruth	IMU-groundtruth
x/m	2.6959	35.0502	25.8619
y/m	2.2211	96.6908	44.9099
z/m	2.2949	77.0708	0.2860

Visual Inertial Odometry Is Disturbed by Continuous Noise, While Other Sensors Operate Normally

In the following **Figures 15–18**, the Fusion trajectory represents the trajectory when all sensors are working normally, the VIO inference trajectory represents the trajectory when the VIO is disabled by noise interference, while **Figure 16** is the trajectories for **Figure 15** t ∈ [200, 210] and t ∈ [200, 300] sections and enlarged to allow a more accurate judgment of the results.

(1) Trajectory comparison in *x*-axis direction.
(2) Comparison of the trajectory in the *y*-axis direction.
(3) Comparison of trajectory in *z*-axis direction.

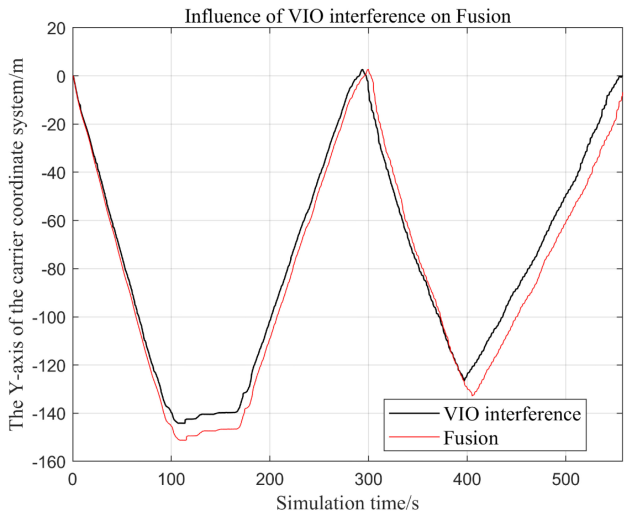


FIGURE 17 | Comparison of the overall trajectory in *y*-axis after VIO failure and when the sensor is working normally.

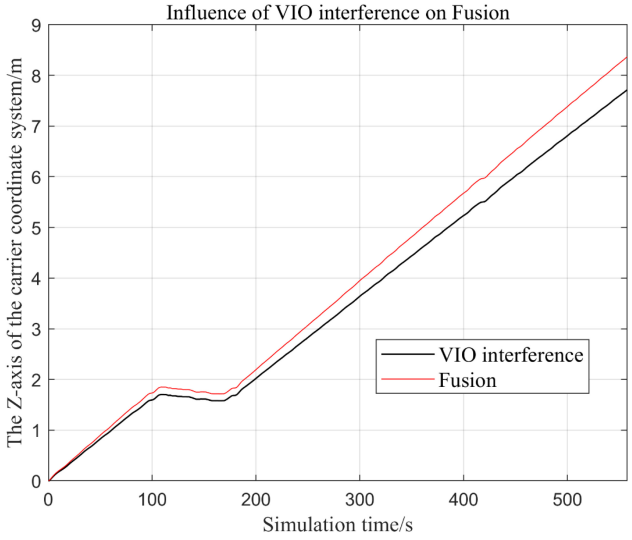


FIGURE 18 | Comparison of the overall trajectory in *z*-axis after VIO failure and when all sensors are working normally.

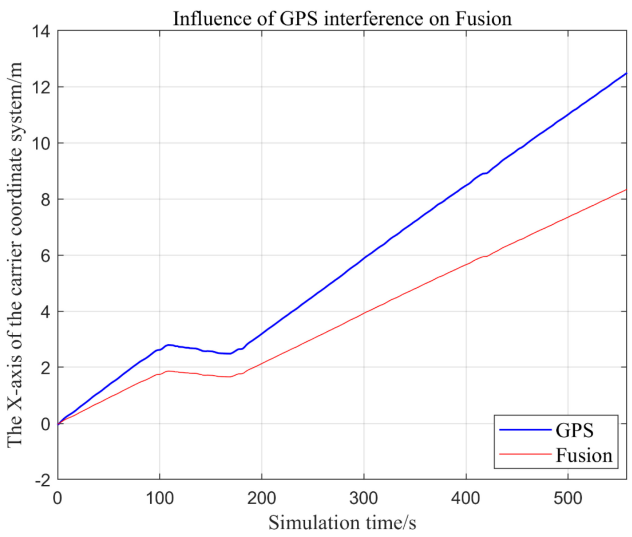


FIGURE 19 | Comparison of overall trajectory in *x*-axis after GPS failure and when all sensors are working normally.

Global Positioning Navigation System Is Continuously Disturbed by Noise, Other Sensors Are Working Normally

From **Figures 19–22**, the Fusion trajectory represents the trajectory when all sensors are working normally, the GPS represents the trajectory when the GPS is disabled by interference, while **Figure 20** are the trajectories of **Figure 19** t ∈ [200, 300] sections and enlarged to allow a more accurate judgment of the results.

(1) Comparison of trajectory in *x*-axis direction.
(2) Comparison of the trajectory in the *y*-axis direction.
(3) Comparison of trajectory in *z*-axis direction.

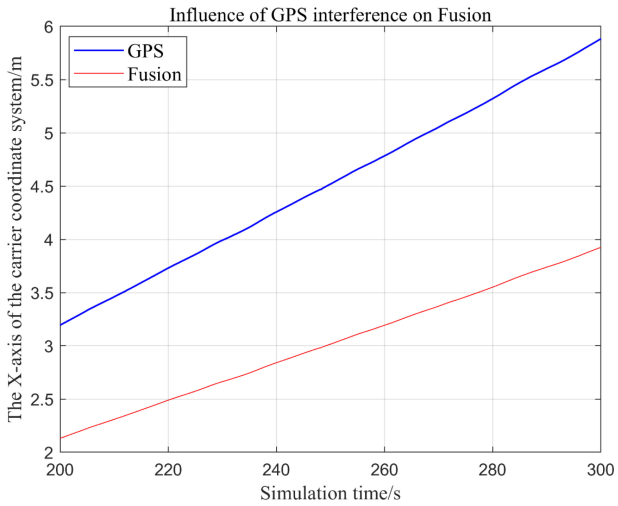


FIGURE 20 | Comparison of the local trajectory in *x*-axis after GPS failure and when all sensors are working normally.

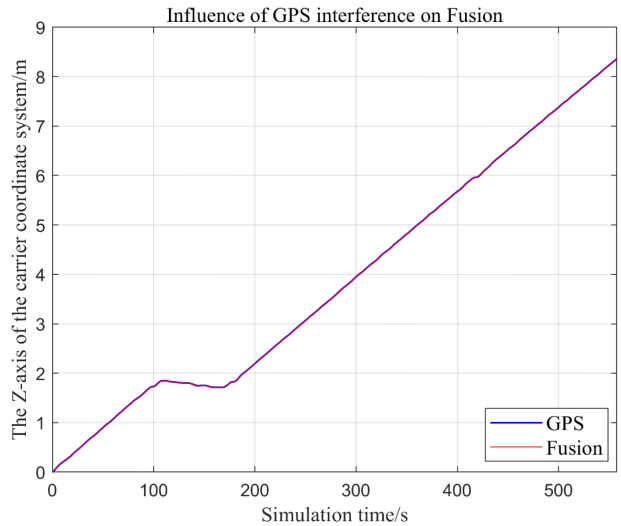


FIGURE 22 | Comparison of the overall trajectory in *z*-axis after GPS failure and when the sensor is working normally.

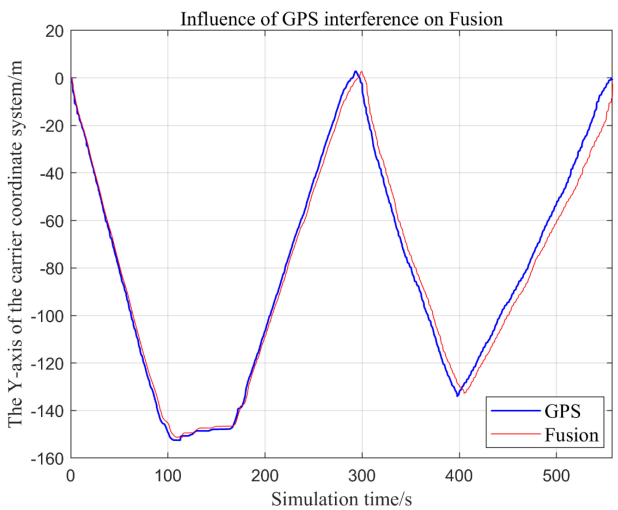


FIGURE 21 | Comparison of the overall trajectory in *y*-axis after GPS failure and when the sensor is working normally.

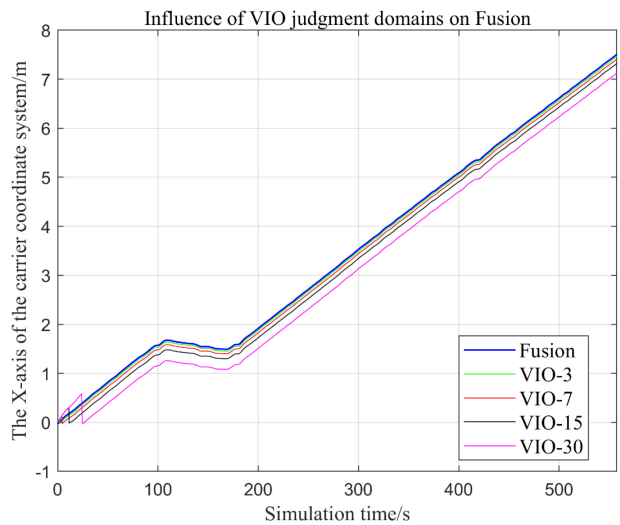


FIGURE 23 | Comparison of the overall trajectory of the output *x*-axis with different thresholds.

Analysis of the Results

In this paper, we disable the sensors GPS and VIO respectively when t ∈ [210, 300], and enlarged the section of figure when t ∈ [200, 300]. In this way, we compare the two output trajectories and find that the robustness of the fusion algorithm proposed in this paper is better.

The above analysis of **Figures 15–22** shows that the GPS or VIO failure does not have much effects on the proposed algorithm results, and it can be thus concluded that the proposed algorithm has excellent robustness ability to remain stable under a continuous disturbance.

Effects of Different Judgment Conditions of Visual Inertial Odometry Algorithm Fusion on the System

Comparison of the Trajectories of the Fusion Algorithm Under Different Judgment Conditions

In the process of sensor fusion, this paper defines a threshold value for VIO sensors and sets judgment conditions based on this threshold value. If the VIO signal greater than this value, it is judged that the VIO data is not suitable for fusion, and the fusion algorithm of the VIO data is terminated.

In this paper, by changing the threshold value and observing the output trajectory, we study the influence of different

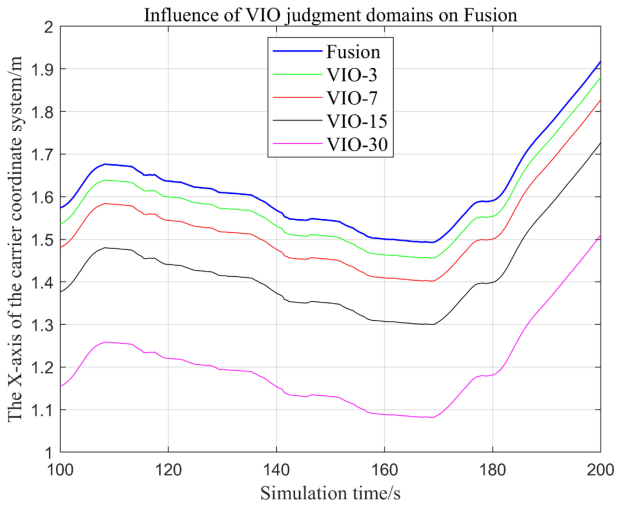


FIGURE 24 | Comparison of output *x*-axis local trajectories under different thresholds.

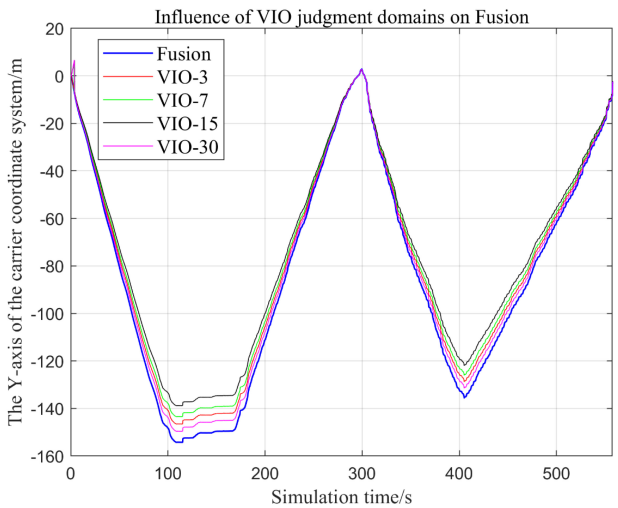


FIGURE 25 | Comparison of the overall trajectory of the output *y*-axis under different thresholds.

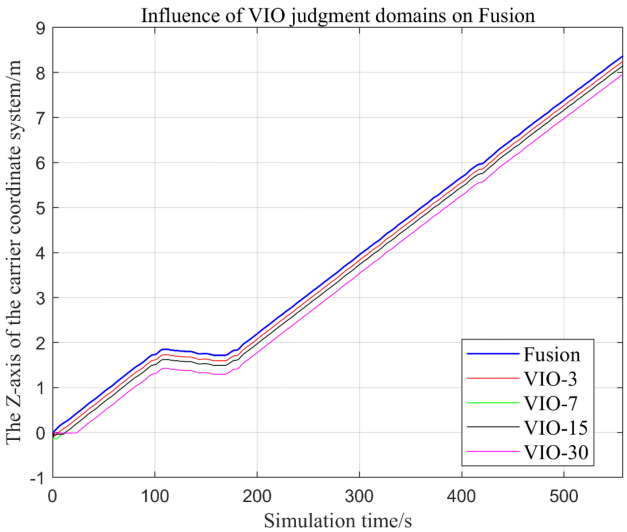


FIGURE 26 | Comparison of the overall trajectory of the output *z*-axis under different thresholds.

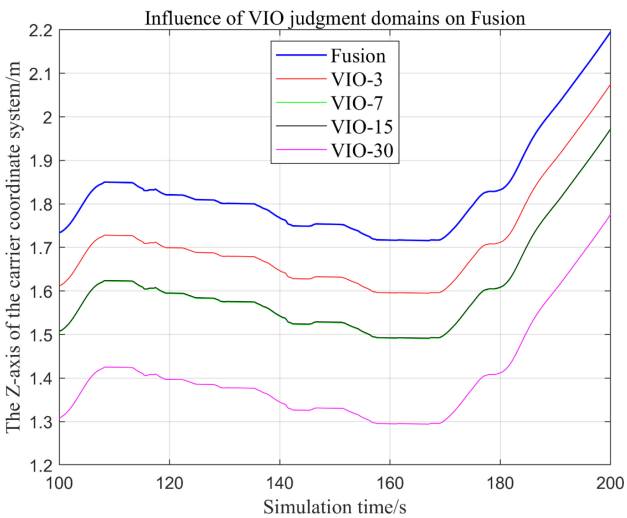


FIGURE 27 | Comparison of output *z*-axis local trajectories under different thresholds.

thresholds judgment conditions on the trajectory results of the proposed fusion algorithm.

In the following **Figures 23–27**, the Fusion trajectory represents the output trajectory when the threshold value is not changed, which is also the Fusion trajectory in all previous simulations with a threshold value of 0.3. VIO-3 represents the output trajectory with a threshold value of 3. VIO-7 represents the output trajectory with a threshold value of 7. VIO-15 represents the output trajectory with a threshold value of 15. And VIO-30 represents the output trajectory with a threshold of 30. **Figures 24**, **27** are the local trajectory results of t ∈ [100,200] sections from **Figures 23**, **26**, respectively.

(1) Comparison of trajectories in the *x*-axis direction.
(2) Comparison of trajectories in the *y*-axis direction.
(3) Comparison of trajectories in *z*-axis direction.

Analysis of Results

In this paper, by changing the defined threshold value of VIO in the proposed algorithm and thus changing the fusion judgment condition, we observe the output trajectory results and analyze the effects of different thresholds on the overall algorithm outputs.

It can be seen from **Figures 23–27** that, as the threshold value increases, the range of judging the VIO data to meet the fusion condition also increases. Based on the comparison between the output trajectories corresponding to different threshold values and the Fusion trajectories, it is observed that the trajectories on the *x*-axis and *z*-axis gradually approach the standard trajectories. It is thus inferred that the accuracy of the proposed fusion

algorithm is improved with the increase of the threshold value within a certain range.

From the above experimental results, it can be concluded that the proposed multi-sensor fusion algorithm has a higher stability compared with traditional VIO algorithms such as MSCKF_VIO and the fusion algorithm of IMU and ODOM fusion algorithm. In addition, it also has excellent robustness. As the working time of the robot increases, the algorithm can still maintain a relatively stable trajectory, make up for the shortcomings of a single VIO, and thus solves the possible target loss and trajectory drift. It should be noted that, although the accuracy of the proposed algorithm has been greatly improved compared with the traditional VIO algorithm, there is still much room for accuracy improvement. Based on the relationship between the accuracy of the algorithm output trajectory and different threshold values, it is meaningful to find an optimal threshold values, in order to stabilize the trajectory errors in a small interval and make the algorithm output trajectory close to the standard trajectory.

CONCLUSION

This paper proposes a loosely coupled EKF MSF algorithm for designing navigation systems. A series of experiments verified that the proposed algorithm has favorable robustness and stability against other methods. The proposed method provide reference significance for tasks such as navigation, localization, and path planning of agricultural robots.

In the future, we will establish a more extensive and complex dataset that are closer to practical applications, in order to further improve the robustness and accuracy of the algorithm under fast motion and more complex random scenarios. It is also interesting to find an optimal threshold values, in order to stabilize the trajectory errors in a small interval and make the algorithm output trajectory close to the standard trajectory.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://www.cifasis-conicet.gov.ar/robot/.

AUTHOR CONTRIBUTIONS

ML and HW designed the research. ML and XF wrote the manuscript. ML, DZ, HW, XF, and WW conducted and analyzed the experiments. WW helped to edit the manuscript. DZ supervised the projected and helped to design the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Foundation Enhancement Project of the State Key Laboratory of Science (2020-JCJQ-ZD-076-00), National Natural Science Foundation of China (51977177), Shaanxi Province Key Research and Development Plan 2021ZDLGY11-04, Basic Research Plan of Natural Science in Shaanxi Province (2020JQ-152), the Fundamental Research Funds for the Central Universities (D5000210763).

REFERENCES

Abidi, M. A., and González, R. C. (1992). *Data Fusion in Robotics and Machine Intelligence*. London: Academic Press Professional, 546.

Atefi, A., Ge, Y., Pitla, S., and Schnable, J. (2021). Robotic technologies for high-throughput plant phenotyping: contemporary reviews and future perspectives. *Front. Plant Sci.* 12:611940. doi: 10.3389/fpls.2021.611940

Benet, B., and Lenain, R. (2017). "Multi-sensor fusion method for crop row tracking and traversability operations," in *Proceedings of the 2017 Conference AXEMA-EURAGENG*, Paris, 10.

Delmerico, J., and Scaramuzza, D. (2018). "A bench-mark comparison of Monocular visual-inertial Odometry algorithms for flying robots," in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, (Piscataway, NJ: IEEE), 2502–2509. doi: 10.1109/ICRA.2018.8460664

Faessler, M., Fontana, F., Forster, C., and Scaramuzza, D. (2015). "Automatic re-initialization and failure recovery for aggressive flight with a Monocular vision-based quadrotor," in *Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA)*, (Piscataway, NJ: IEEE), 1722–1729. doi: 10.1109/ICRA.2015.7139420

Faessler, M., Fontana, F., Forster, C., Mueggler, E., Pizzoli, M., and Scaramuzza, D. (2016). Autonomous, vision-based flight and live dense 3D mapping with a quadrotor micro aerial vehicle. *J. Field Robot.* 33, 431–450. doi: 10.1002/rob.21581

Gao, X., Zhang, T., Liu, Y., and Yan, Q. (2017). *14 Lectures on Visual Slam*. Beijing: Electronic Industry Press, 141–161.

Hall, D. L., and Llinas, J. (1997). An introduction to multisensory data fusion. *Proc. IEEE* 85, 6–23. doi: 10.1109/5.554205

Han, J., Park, C., Kwon, J. H., Lee, J., Kim, T. S., and Jang, Y. Y. (2020). Performance evaluation of autonomous driving control algorithm for a crawler-type agricultural vehicle based on low-cost multi-sensor fusion positioning. *Appl. Sci.* 10:4667. doi: 10.3390/app10134667

Hu, H., and Yan, H. (2018). "Multi-sensor data fusion algorithm of wisdom agriculture based on fusion set," in *Proceedings of the 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, (Piscataway, NJ: IEEE), 121–124. doi: 10.1109/ICVRIS.2018.00037

Inam, U., Xin, S., Xuewu, Z., and Dongmin, C. (2020). Simultaneous localization and mapping based on Kalman filter and extended Kalman filter. *Wirel. Commun. Mob. Comput.* 2020:2138643. doi: 10.1155/2020/2138643

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* 82, 35–45. doi: 10.1115/1.3662552

Katsigiannis, P., Misopolinos, L., Liakopoulos, V., Alexandridis, T. K., and Zalidis, G. (2016). "An autonomous multi-sensor UAV system for reduced-input precision agriculture applications," in *Proceedings of the 2016 24th Mediterranean Conference on Control and Automation (MED)*, (Piscataway, NJ: IEEE), 60–64. doi: 10.1109/MED.2016.7535938

Lee, X. Y., and Gao, G. W. (2019). "An improved CI EKF data fusion algorithm for multi-sensor time-delay system," in *Proceedings of the 2019 5th International Conference on Applied Materials and Manufacturing Technology*, Singapore, 1660–1665. doi: 10.26914/c.cnkihy.2019.075301

Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., and Furgale, P. T. (2015). Keyframe-based visual-inertial ODOMetry using nonlinear optimization. *Int. J. Robot. Res.* 34, 314–334. doi: 10.1177/0278364914554813

Liu, P. Z., Bi, S. S., Zang, G. S., Wang, W., Gao, Y., and Deng, Z. (2011). "Obstacle avoidance system for agricultural robots based on multi-sensor information fusion," in *Proceedings of the 2011 International Conference on Computer Science and Network Technology*, (Piscataway, NJ: IEEE), 1181–1185.

Mourikis, A. I., and Roumeliotis, S. I. (2007). "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, (Piscataway, NJ: IEEE), 3565–3572. doi: 10.1109/ROBOT.2007.364024

Noguchi, N., Reid, J. F., Will, J., Benson, E. R., and Stombaugh, T. S. (1998). "Vehicle automation system based on multi-sensor integration," in *Proceedings of the 1998 Annual International Meeting,* Paper No.983111, (St. Joseph, MI: ASAE), 49085–49659.

Ollis, M., and Stentz, A. (1997). "Vision-Based perception for an autonomous harvester," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vol. 3, (Piscataway, NJ: IEEE), 1838–1844.

Pire, T., Mujica, M., Civera, J., and Kofman, E. (2019). The Rosario Dataset: multi-sensor data for localization and mapping in agricultural environments. *Int. J. Robot. Res.* 38, 633–641. doi: 10.1177/0278364919841437

Qin, Z. W., Wang, W. W., Dammer, K. H., et al. (2021). A machine learning approach for real time low-cost detection of palm trees as target for precise crop protection. *Front. Plant Sci.* 12:753603. doi: 10.3389/fpls.2021.753603

Reina, G., and Messina, A. (2019). Vehicle dynamics estimation via augmented extended filtering. *Measurement* 133, 383–395. doi: 10.1016/j.measurement.2018.10.030

Sastry, V. (1971). Decomposition of the extended Kalman filter. *IEEE Trans. Automat. Contr.* 16, 260–261. doi: 10.1371/journal.pone.0209836

Scaramuzza, D., Achtelik, M. C., Doitsidis, L., Fraundorfer, F., Kosmatopoulos, E. B., Martinelli, A., et al. (2014). Vision-controlled micro flying robots: from system design to autonomous navigation and mapping in GPS-denied environments. *IEEE Robot. Automat. Mag.* 21, 26–40. doi: 10.1109/MRA.2014.2322295

Sharma, S., and Borse, R. (2016). "Automatic agriculture spraying robot with smart decision making," in *Intelligent Systems Technologies and Applications 2016. ISTA 2016. Advances in Intelligent Systems and Computing Book Series (AISC)*, Vol. 530, eds J. Corchado Rodriguez, S. Mitra, S. Thampi, and E. S. El-Alfy (Cham: Springer). doi: 10.1007/978-3-319-47952-1_60

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.* 11:510. doi: 10.3389/fpls.2020

Tong, Q., Li, P., and Shen, S. (2018). VINS-MONO: a robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* 34, 1004–1020. doi: 10.1109/TRO.2018.2853729

Varshney, P. K. (2000). "Multisensor data fusion," in *Intelligent Problem Solving. Methodologies and Approaches. IEA/AIE 2000. Lecture Notes in Computer Science*, eds R. Logananthara, G. Palm, and M. Ali (Berlin: Springer), 245–253. doi: 10.1007/3-540-45049-1_1

Velasquez, A. E., Higuti, V. A., Gasparino, M. V., Sivakumar, A. N., Becker, M., and Chowdhary, G. (2021). Multi-sensor fusion based robust row following for compact agricultural robots. Computer Science, Robotics. *arXiv* [preprint]. arXiv: 2106.15029,

Viacheslav, A. I., Rossel, R. A. V., Sudduth, K. A., and Lammers, P. S. (2011). "Sensor fusion for precision agriculture," in *Sensor Fusion Foundation and Applications*, ed. C. Thomas (Rijeka: InTech), 27–40. doi: 10.5772/19983

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 26 April 2022
doi: 10.3389/fpls.2022.870181

Cotton Yield Estimation From Aerial Imagery Using Machine Learning Approaches

Javier Rodriguez-Sanchez[1], *Changying Li*[1,2*] *and Andrew H. Paterson*[2,3]

[1] Bio-Sensing and Instrumentation Laboratory, College of Engineering, The University of Georgia, Athens, GA, United States, [2] Phenomics and Plant Robotics Center, The University of Georgia, Athens, GA, United States, [3] Plant Genome Mapping Laboratory, The University of Georgia, Athens, GA, United States

Estimation of cotton yield before harvest offers many benefits to breeding programs, researchers and producers. Remote sensing enables efficient and consistent estimation of cotton yields, as opposed to traditional field measurements and surveys. The overall goal of this study was to develop a data processing pipeline to perform fast and accurate pre-harvest yield predictions of cotton breeding fields from aerial imagery using machine learning techniques. By using only a single plot image extracted from an orthomosaic map, a Support Vector Machine (SVM) classifier with four selected features was trained to identify the cotton pixels present in each plot image. The SVM classifier achieved an accuracy of 89%, a precision of 86%, a recall of 75%, and an F1-score of 80% at recognizing cotton pixels. After performing morphological image processing operations and applying a connected components algorithm, the classified cotton pixels were clustered to predict the number of cotton bolls at the plot level. Our model fitted the ground truth counts with an R^2 value of 0.93, a normalized root mean squared error of 0.07, and a mean absolute percentage error of 13.7%. This study demonstrates that aerial imagery with machine learning techniques can be a reliable, efficient, and effective tool for pre-harvest cotton yield prediction.

Keywords: cotton yield estimation, machine learning, UAS, SVM, remote sensing

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Aichen Wang,
Jiangsu University, China
Patricia Terletzky-Gese,
Utah State University, United States
Phuong Dao,
University of Wisconsin-Madison, United States

***Correspondence:**
Changying Li
cyli@uga.edu

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *06 February 2022*
Accepted: *21 March 2022*
Published: *26 April 2022*

Citation:
Rodriguez-Sanchez J, Li C and Paterson AH (2022) Cotton Yield Estimation From Aerial Imagery Using Machine Learning Approaches. Front. Plant Sci. 13:870181. doi: 10.3389/fpls.2022.870181

1. INTRODUCTION

Cotton is a major industrial crop in the United States (U.S.), especially in the southern and western U.S. Cotton fiber is one of the principal natural textile fibers worldwide (Townsend and Sette, 2016), and the U.S. is the third leading cotton producer with an expected production of 22.5 million bales for 2019/20, just after China (27.5 million bales) and India (27 million bales). Cotton is a soft staple fiber that grows from the surface of seeds, enclosed in pods known as bolls. Primary components of economic yield, cotton boll number and boll weight are agronomic traits that help to define cotton crop performance in its last stages of growth. These traits can be used as indicators of fiber production, which ultimately play a key role in breeding programs and may also provide valuable information for farmers to plan hedging strategies.

Lint yield is one of the most important criteria for selecting new lines in breeding programs (Bourland and Myers, 2015), but it is costly to obtain reliable data. Visual estimation of yield performance is often used by cotton breeders to select promising cotton genotypes, but it can be challenging. Morphological characteristics of cotton plants such as general shape, branch density,

or leaf area change during the growth cycle of the crop and they may mislead visual ratings of yield (Bourland and Myers, 2015). Moreover, boll size can vary by year, breeding line, and position on each fruiting branch, which can make it difficult to standardize visual cotton yield quantification methods. Physical harvesting of the bolls, either manually or by using mechanical pickers, to reduce the estimation bias is labor intensive and time-consuming, limiting the number of plots that can be quantified (Bowman et al., 2004). Thus, the development of tools for effectively automating plant phenotyping tasks is of great potential value for breeding programs.

In recent years, the applications of unmanned aerial systems (UAS) in agriculture have grown rapidly and have transformed modern farming. UAS are relatively inexpensive and can be equipped with a variety of sensors, which makes them a valuable tool for large crop field monitoring. These systems can be programmed easily to navigate pre-defined paths with a specific velocity while retaining a specific distance from the crop. This means that they can be used to collect data remotely from the field at optimal resolutions quickly and easily. UAS surveying has been widely used for monitoring different crops (Barbedo, 2019), but only a few studies have addressed the use of these systems to estimate cotton yield, and only two have investigated their use for cotton genotype selection. The methodologies used to estimate cotton yield from UAS imagery can be classified in two main groups: approaches based on the use of only 3-channel RGB (Red, Green, Blue) color images, and approaches that utilize a combination of different sensor data to indirectly calculate lint yield.

To estimate cotton yield from RGB imagery, one of the main techniques is color thresholding segmentation, which has been applied either to a single color channel or to multiple channels at the same time. For instance, Dodge (2019) applied a global thresholding method (using a fixed threshold value) to the B channel alone on RGB aerial images to isolate cotton-related pixels. Their methodology achieved an R^2 of 0.817 for the first year's experimental data. However, this relationship was not consistent, and they needed to include additional postprocessing to improve their model generalization for the next year's data (R^2 = 0.736). Yeom et al. (2018) analyzed the spatial and spectral characteristics of open cotton bolls on RGB images during the harvest period. They established a global automatic threshold based on Otsu's method to separate cotton bolls from other non-target objects. They achieved R^2 values of 0.63–0.65 at estimating yield using the cotton boll area as the input variable. However, they assumed that cotton bolls have higher spectral values than the other elements of the crop, which can be a limiting factor with changing illumination conditions when the range of image intensities of the color channels for the cotton bolls can resemble other crop elements. Huang et al. (2016) found that a global RGB threshold could not extract the cotton pixels from the images accurately because the range of image intensities in the R, G, and B channels of the cotton bolls overlapped with that of the soil and other crop elements. Alternatively, they proposed the application of the thresholding technique on Laplacian images obtained from the divergence of the gradient of each image with respect to pixel intensity. They were able to establish a ratio—cotton unit coverage (CUC)—of the number of cotton boll pixels detected to the number of pixels in a particular area. They achieved their best results (R^2 = 0.83) after introducing additional postprocessing steps to detect and remove poorly illuminated plot images because their method was affected by shadowing and changing illumination conditions.

Additional approaches based only on RGB images have also been proposed. Maja et al. (2016) estimated cotton yield of small field plots from a cotton breeding program using K-means clustering algorithm with 4 classes. They clustered the cotton pixels on the image based on their color and found a linear relationship (R^2 = 0.782) between the ratio of cotton pixels with respect to the total image area and the actual yield. However, they needed to introduce a fixed cluster size constraint to avoid large clusters and reduce misclassification of highly reflective areas of the scene such as the bare soil. This additional constraint can limit the generalization of their methodology for highly dense cotton crops where the cotton bolls tend to form large groups. Chu et al. (2016) estimated additional crop information (plant height and canopy cover) from aerial RGB images and were able to model cotton yield (R^2 = 0.529) before maturation and boll opening. However, their study of yield was limited to the stages of the crop before defoliation such that the canopy cover could be computed correctly.

In addition to RGB cameras for lint yield estimation, multispectral and thermal cameras have been used. Huang et al. (2013) used various vegetation indices obtained from multispectral aerial images in conjunction with soil properties to estimate yield variation. Their model based on the ratio of vegetation index (RVI) and soil electrical conductivity (EC) measurements performed well for non-irrigated fields (R^2 = 0.718) but was unable to estimate yield accurately for irrigated fields (R^2 = 0.033). Feng et al. (2020) modeled cotton yield using multiple features derived from RGB, multispectral, and thermal cameras. They applied a global threshold in all three channels R, G, and NIR (near infrared radiation) to discriminate open cotton bolls from the soil and leaves. They found that the ratio of the number of cotton pixels to the overall number of pixels in a specific area of a multispectral image (cotton fiber index, CFI) could be used to estimate yield at the pre-harvest stage (R^2 = 0.90). Moreover, by using a combination of plant height, CFI, canopy temperature and the a* component of the CIELAB color space, they obtained an even better result (R^2 = 0.94). However, this sophisticated approach required the simultaneous use of color, multispectral and thermal cameras, which is costly and may require more computation capacity, time, and labor for data collection and processing.

Methods based on machine learning (ML) techniques have been explored recently. Support vector machines (SVM) are one of the most widely used machine learning algorithms for supervised data classification and regression analysis. Based on statistical learning, SVMs aim to identify a decision boundary to partition data in a high-dimension feature space into two sets. This decision hyperplane can be then used for data classification or regression analysis. There are two basic SVM formulations to perform these tasks. For data classification, support vector classification (SVC) models (Cortes and Vapnik, 1995) try to

find the hyperplane to separate the input data belonging to two different classes with the maximum margin. The learning process for an SVC aims to maximize that margin and minimize classification errors between the two classes. These classification models return for each input data a class label and its probability of belonging to each class. The second type of SVM is for regression problems. Support vector regression (SVR) is a regression function that can predict dependent variable by using independent variables as continuous values instead of class labels. The SVR works with the similar principle as SVM: to find the hyperplane that best fits the data inside a decision boundary delimited by a predefined error margin (Drucker et al., 1996; Vapnik et al., 1997). The models can be optimized using the regularization parameter C and the margin of tolerance ϵ. During the last decade, SVMs and its variants have been successfully applied in agricultural remote sensing for crop classification (Song et al., 2014; Liu and Whitty, 2015) and plant disease identification (Rumpf et al., 2010; Garcia-Ruiz et al., 2013; Raza et al., 2015). For cotton crops, SVMs have been successfully used to identify cotton flowers from multispectral imagery (Xu et al., 2019) and for budding rate monitoring from aerial RGB imagery (Xia et al., 2019). Regarding cotton yield estimation, other ML methods have also been used. By using RGB and multispectral aerial imagery, Ashapure et al. (2020) implemented an artificial neural network (ANN) to estimate cotton yield throughout the season based on crop canopy attributes. They were able to predict cotton yield at early stages of the crop with an average $R^2 = 0.861$ using features that included canopy information, multispectral vegetation indices, cotton boll information (obtained by using a color-based threshold method, Jung et al., 2018), and crop irrigation status. This advanced methodology required the use of color and multispectral imaging systems, which increases platform costs and would entail additional time and labor for data collection and processing.

Methods for cotton yield estimation based on remote sensing techniques have various limitations. Traditional image processing techniques, such as thresholding, are not flexible enough to adapt to variable conditions present in the scene, resulting in relatively low performance. Some studies rely upon additional sensors to retrieve new features, which ultimately increases platform cost and usage and processing complexity. In this article, we present a simple and easy-to-implement methodology for estimating pre-harvest cotton yield at the plot level. By using RGB airborne imagery and a SVM algorithm, this new methodology could benefit cotton breeders by allowing them to acquire accurate information about different crop plots for their selection experiments in a timely manner. Moreover, it also may provide farmers an inexpensive, quick, and precise estimation of the yield in their cotton fields before harvesting. The specific objectives of this study were to (1) implement a support vector machine model to classify cotton boll pixels in aerial images; (2) develop a predictor to estimate the number of cotton bolls at the plot level; (3) evaluate the performance of the proposed method against ground truth yield measurements; and (4) apply the proposed method to identify differences between genotypes in a cotton breeding trial.

2. MATERIALS AND METHODS

2.1. Experimental Field

The field under study was located at the Iron Horse Farm (IHF) in Greene County, Georgia, U.S. (33°43′01.3″N 83°18′29.1″W) (**Figure 1**). The dimensions of the field were 200 × 12m (length × width), and it comprised a total of 488 plots. These plots were arranged into 10 rows with 46 plots per row and 1 additional row of 28 plots. A total of 220 different cotton genotypes were planted in plots of approximately 3-meters wide, with a final plant density dependent on the germination rate of the seeds. Six of the genotypes were commercial cultivars (TAM94L25, Acala1517-98, UA48, FM832, DeltaPine393, and GA230) with 10 replications per cultivar. The rest were 214 breeding lines from 10 different cotton populations (J, K, L, M, N, O, P, Q, R, S) with two replicates.

2.2. Data Collection

2.2.1. Aerial Imagery

Original RGB color images were captured on February 1, 2020 on a single flight using a quadcopter DJI Matrice M100 (Shenzhen DJI Sciences and Technologies Ltd., Shenzhen, China), equipped with a Lumix G7 digital single-lens reflex (DSLR) camera (Panasonic Corporation, Osaka, Japan). This camera has a 17.3 × 13 mm CMOS image sensor with 16.0 megapixels (4592 × 3448 pixels) resolution and stores captured images using the sRGB color space. The camera was mounted on the bottom of the drone using a custom 3D-printed bracket, which ensured that the camera lens was aligned to a 90 degree angle relative to the ground. **Figure 2** shows the system used for data collection. The flight was controlled internally by the M100's N1 flight controller and was carried out at a height of 15 m above ground level, and at a speed of 1.9 m/s. With this configuration, the ground pixel size was 0.26 cm/pixel. The camera was configured automatically according to the light conditions of the field. Different white balance configurations were tested for color balancing before the flight, and the "Auto White Balance" compensation (AWB) was found adequate for the weather conditions on the collection day. A Manifold onboard computer (Shenzhen DJI Sciences and Technologies Ltd., Shenzhen, China) was in charge of triggering the camera at a constant rate of 1 frame per second. The forward overlap between images in the same flight line was configured to 80%, while the side-by-side overlap between adjacent flight lines was set to 60%. A total of 447 images were collected during the flight.

A set of 12 ground control points (GCP) with circular patterns were generated using Agisoft Metashape software (Metashape Professional 1.5.5, Agisoft LLC, Russia) and deployed along the field's border for geo-referencing the UAS images (**Figure 3**). The patterns were cut off from matte black adhesive-backed vinyl sheet and pasted onto a white acrylic 60 cm square sheet. The GCPs were surveyed in the field using an RTK-GPS system.

2.2.2. Ground Truth Data

A digital method was used to provide the ground truth data to evaluate the remote sensing predictions. Since manual harvesting of cotton bolls is time consuming, labor intensive, and

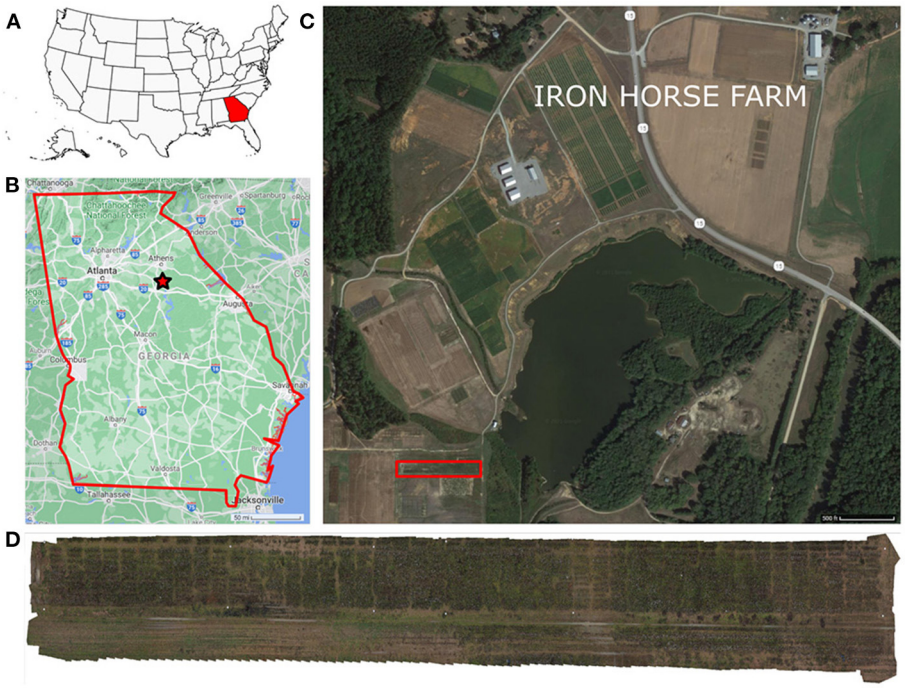


FIGURE 1 | Experimental field location. The experiment was conducted in the Iron Horse Farm, Greene County, GA, U.S. **(A)** Location map of Georgia, marked in red, in the U.S., **(B)** General map of Georgia, the experiment location is marked with a red star, **(C)** Iron Horse Farm aerial view (Greene County, GA), the specific location of the experimental field is delimited by a red rectangle, **(D)** Field layout.

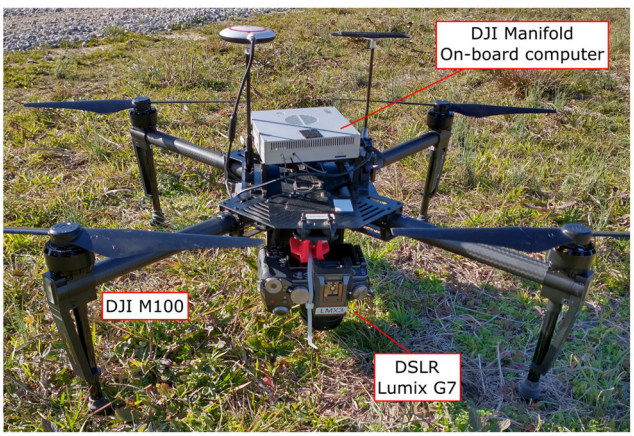


FIGURE 2 | Equipment used for image collection. Unmanned Aerial Vehicle, DJI Matrice M100 Pro, equipped with a Manifold onboard computer and a Panasonic Lumix DMC-G7 DSLR camera as imaging device.

destructive, we employed high resolution 3D point clouds and a virtual reality (VR) annotation tool to count number of bolls digitally. The field was scanned on February 2nd using terrestrial laser scanning (TLS) techniques. A FARO Focus S70 3D laser scanner (FARO Technologies Inc., Florida, U.S.) was used to collect high resolution 3D point cloud data (PCD) from multiple locations through the field. The scanner was configured to $\frac{1}{2}$ resolution, and 2× quality. With these parameters a full scan can collect up to 174.8 megapoints with a point distance of 3.1 mm in a scan distance of 10 meters. The LiDAR data was captured from the ground at a distance ranging between 1 and 2 m from the plants, to enable the visual identification and counting of the cotton bolls. Individual scans were registered as a single PCD using FARO SCENE 2019 software (FARO Technologies Inc., Florida, U.S.). A set of 10 coded planar markers obtained from SCENE software were deployed around the field to facilitate the coregistration of multiple PCDs. SCENE was configured to automatically detect the registration markers and align the scans. After registration, individual plots were manually segmented and extracted from the 3D reconstructed PCD as .PTS files using SCENE *clipping box* tool. Each individual plot was then processed using CloudCompare software (version 2.11.2). A statistical outlier removal using 2 points for mean distance estimation and 1.0 as the standard deviation multiplier threshold was applied to the individual point clouds to reduce noise and clean spurious points. From the 488 plots that composed the field, a set of 45 plots (**Figure 3**) were selected for digital ground truth counting. A VR annotation tool developed by the Virtual Experience Laboratory at the University of Georgia was leveraged to count the cotton bolls from the clean PCDs using an Oculus Quest 2 VR set (Facebook Technologies Inc., California, U.S.) and a desktop computer. More details of this VR tool will be covered in a future manuscript. These counts were considered as the ground truth measurements (GT_{VR}) for further analysis.

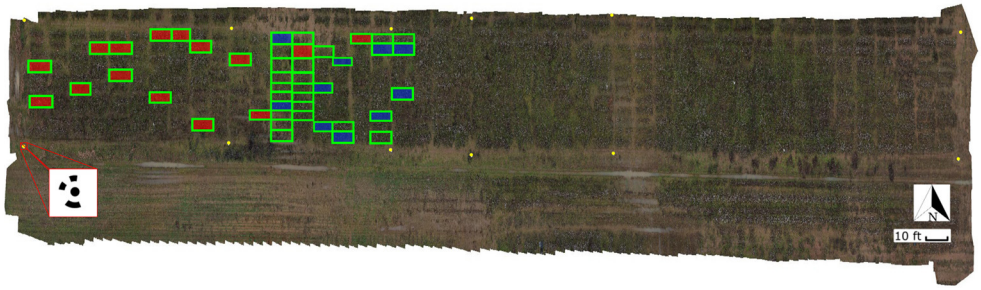


FIGURE 3 | General orthomosaic map of the experimental field. Ground Control Points (GCP) used for image geo-referencing are identified as yellow circles; a total of 12 GCPs were used around the field. The red square at the bottom-left corner shows a zoom-in view of one of the 60 × 60 cm square GCPs with the 12-bit circular coded pattern. Ground truth plots measured on February 2, 2020 using digital approaches are identified with green boxes (45 plots). Plots harvested on February 3, 2020 are identified with boxes filled in red (15 plots); ground truth plots harvested on February 27, 2020 are identified with boxes filled in blue (10 plots).

TABLE 1 | Ground truth data summary.

	Samples no.	Min (boll number)	Max (boll number)	Mean (boll number)
Digital ground truth	45	14	362	160.933
Manual ground truth	25	61	367	184.680

Manual ground truth counts include the cotton boll number measured by destructive sampling of 25 plots randomly selected from the field. Digital ground truth counts include the cotton boll number of 45 ground truth plots measured using digital approaches.

The plots used as the ground truth included a representative subset of the plots in the field: from plots with a small number of cotton bolls (<20 cotton bolls) to highly dense cotton plots (>350 cotton bolls). A summary of the ground truth values for both the manual and digital sampling is presented in **Table 1**. To calibrate this approach, the digital ground truth counts were regressed against the actual number of cotton bolls in a subset of manually harvested plots. The manual ground truth subset was composed of 25 plots that were randomly selected from the 45 plots in the digital ground truth set. In these plots, only the open cotton bolls were harvested, counted, and weighted manually. These manual ground truth measurements (GT_{manual}) were performed in two different batches: 15 plots were harvested on February 3rd, 2020 and 10 plots were harvested on February 27th, 2020 (**Figure 3**). A strong linear relationship ($R^2 = 0.996$) was found between GT_{VR} and GT_{manual} for the 25 manually harvested plots.

2.3. Data Processing Pipeline

The data processing pipeline for cotton yield estimation presented in this article (**Figure 4**) involved four main steps: (A) generation of an orthomosaic map of the entire field from the aerial images collected, (B) individual plot images extraction and pre-processing using image processing techniques, (C) development of an image pixel classifier based on SVM for cotton pixels segmentation at the plot level, and (D) cotton boll number estimation for each individual plot.

2.3.1. Orthomosaic Map Generation and Individual Plot Image Extraction

An orthomosaic map was created from the RGB images using Agisoft Metashape software. A generic pair preselection with high accuracy setting was selected for photo alignment on Metashape software. By using the "*detect markers*" tool, all GCP markers were identified to georeference the images.

After applying a mild depth filtering and enabling interpolation, a digital elevation model (DEM) was generated using the dense point cloud from the estimated camera positions. Finally, using the DEM and mosaic as the blending mode, the orthomosaic map was obtained. To extract plot images from the orthomosaic map we used the open-source geographic information system (GIS) software Quantum GIS, version 3.8.2-Zanzibar (Open Source Geospatial Foundation, Beaverton, OR, U.S.). A vector layer with the boundaries of each plot was created manually and then clipped to the orthomosaic map to obtain an individual GeoTIF image file for each plot.

2.3.2. Individual Plot Image Preprocessing

To reduce computation time and speed up the processing of individual plot images, preprocessing was implemented. Each individual plot image contains around 500,000 pixels. However, just a small portion of these pixels are meaningful for yield estimation, i.e., cotton-related pixels. As a general example, an image of a typical cotton crop plot will include plant leaves, weeds, and other plant matter (green to yellow color pixels); soil and other mineral matter with near-neutral hues (gray, brown, and tan color pixels); branches and other woody elements (light brown color pixels); and cotton pixels (shades of white color pixels). Hence, using traditional image processing techniques, all vegetation and soil-related pixels from the images could be potentially removed prior to pixel classification. A modified version of the Excess Green minus Excess Red Index (*ExG-ExR*) (Meyer and Neto, 2008) provided good results for removing vegetation pixels. To compute the modified index $ExGR_{mod}$ index for the cotton plots, the following equation was used:

$$ExGR_{mod} = 2.5 \times G_{norm} - 3 \times R_{norm} - B_{norm} \quad (1)$$

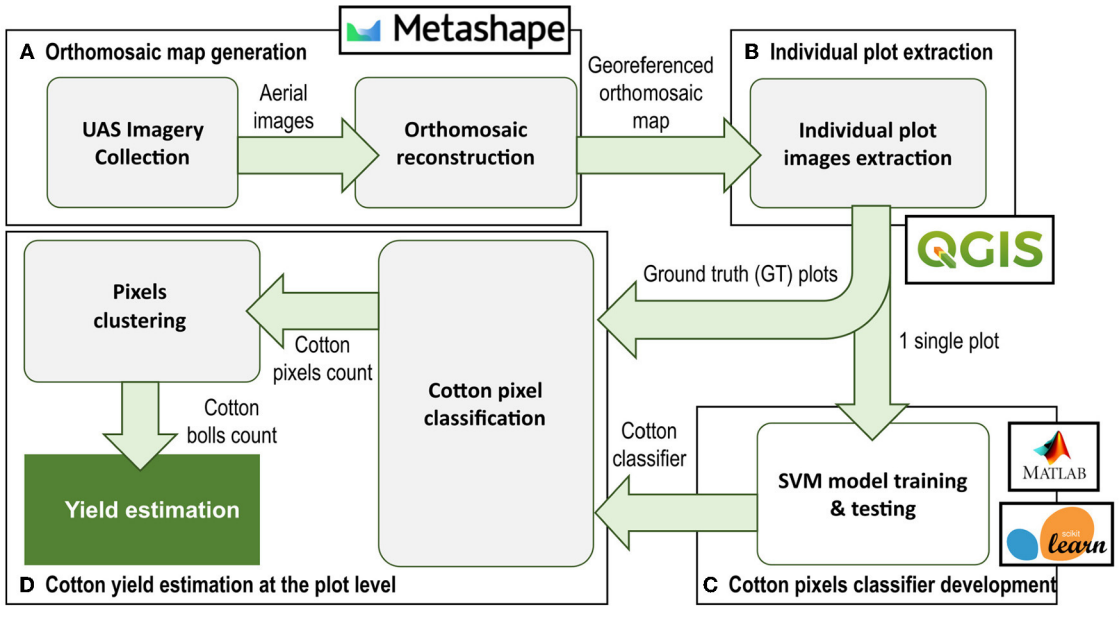


FIGURE 4 | Data processing pipeline. The process pipeline includes 4 main steps: **(A)** Orthomosaic map generation, **(B)** Individual plot extraction, **(C)** Image pixel classifier development, and **(D)** Cotton yield estimation at plot level. Green arrows indicate the flow of data/information between processes.

where R_{norm}, G_{norm}, and B_{norm} are the normalized R, G, and B color channel values respectively that were computed as in Meyer and Neto (2008).

Similarly, a new index that we called $Soil_{Idx}$ based on the CIELAB color space was found useful to remove soil pixels without having any visible impact on the pixels associated to the cotton bolls. To compute this index the following equation was used:

$$Soil_{Idx} = 0.5 \times L_{norm} - 2 \times a_{norm} + b_{norm} \quad (2)$$

where L_{norm}, a_{norm}, and b_{norm} are respectively the normalized L^*, a^*, and b^* components of the CIELAB color space.

2.3.3. SVM Classification Model Development

Cotton color is significantly different from most other elements in the field. Hence, pixel color can be used intuitively as a descriptor for cotton pixel segmentation. However, after a preliminary analysis of the RGB color component values of the image pixels, we found that some branches and woody elements in the background were almost indistinguishable from the cotton boll pixels, mainly because of shades and other light-blocking effects. This suggests that RGB color space information alone was not a robust enough descriptor to properly extract the cotton pixels, as previous studies have noted. Other color spaces, in particular HSV and CIELAB color models, increase invariance with respect to luminosity and lighting changes and are more robust than the RGB color space in relation to the presence of shadows (Hdioud et al., 2018). In this study, we applied an SVM model to classify image pixels using RGB, HSV, and CIELAB color spaces information. This information was used as feature descriptors to discriminate between cotton boll pixels and the rest of background pixels.

The SVM model was developed using the Scikit-learn library (Pedregosa et al., 2011) on the Jupyter Notebook interactive computing platform, version 6.1.4. To reduce the annotation burden, only a single plot image was used to extract features and create the dataset for model training. Initially, an 11-dimensional feature vector was extracted from each pixel. These vectors contained the location of the point in the image (row, col); and the values of the RGB, HSV and CIELAB color space components of the point (R, G, B, H, S, V, L*, a*, b*) for each pixel in the image. The Matlab Image Labeler app (The Math Works Inc., Natick, MA, U.S.) was used to annotate the image. This annotation tool enables the user to interactively draw pixel ROIs to label the boundaries of the visible cotton bolls to classify every image pixel into one of two target classes: cotton and non-cotton. The class of each pixel—1 for cotton pixels, and 0 for non-cotton pixels—was added to the features vector as the target column. To minimize the complexity of the model, a dimension reduction step was introduced to identify the best set of features. A recursive feature elimination (RFE) algorithm was applied for best features selection. The resulting dataset was divided into training and validation subsets with a ratio of 4:1. For model training, a radial basis function (RBF) kernel was used, the hyperparameter C was configured to be 1.0, and the hyperparameter γ was selected as "scale".

To evaluate the performance of the SVM model for cotton pixels classification, accuracy, precision, recall, and Type I and Type II error rates were calculated. In addition, to provide a more comparable metric with other similar studies, the F1 score metric was also computed. The accuracy can be defined as the percentage

of correct predicted pixels for the total number of pixels analyzed and can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3)$$

where TP and TN—true positives and true negatives, respectively—are the number of pixels correctly classified for each class, and FP and FN—false positives and false negatives, respectively—are the number of misclassified pixels.

The precision metric measured the proportion of pixels classified as cotton pixels that were classified correctly. This metric accounts for the ability of the classifier to not label a pixel that is not cotton as cotton. In contrast, the Type I error rate indicates the probability of misclassifying an non-cotton object as cotton using the classifier. The precision and the Type I error rate can be calculated as follows:

$$precision = \frac{TP}{TP + FP} \times 100 \quad (4)$$

$$Type\ I\ error\ rate = \frac{FP}{TP + FP} \times 100 \quad (5)$$

The recall measured the proportion of cotton pixels that were classified correctly by the SVM model among all the actual cotton pixels in the image. This metric describes the ability of the classifier to find all cotton pixels. The Type II rate, in contrast, indicates the probability of misclassifying a cotton pixel as a non-cotton pixel. These metrics can be calculated as follows:

$$recall = \frac{TP}{TP + TN} \times 100 \quad (6)$$

$$Type\ II\ error\ rate = \frac{TN}{TP + FP} \times 100 \quad (7)$$

Finally, F1 score, as a function of precision and recall, conveyed the balance between those two metrics by taking their weighted average. It can be calculated using the following equation:

$$F1\text{-}Score = 2 \times \frac{precision \times recall}{precision + recall} \times 100 \quad (8)$$

All these metrics ranged from 0% to 100%, 100% being related to the best performance.

2.3.4. Yield Prediction Model Development and Evaluation Metrics

The developed SVM classifier was used to classify the cotton pixels presented on the unseen 45 plot images corresponding to the ground truth plots. After each image pixel was classified as a cotton or non-cotton point, a morphological erosion operation using a 3 × 3 elliptic structuring element, followed by a morphological dilation operation with a 5 × 5 rectangular structuring element were applied to eliminate noisy points and reduce the effect of pixels misclassification. Then, connected components labeling was applied to the binary image to count number of bolls. The connected components were computed using an 8-way pixel connectivity, where pixels are considered connected if they share any of the pixels that compose their respective Moore Neighborhood. These post-processing operations were performed using the OpenCV library (Bradski, 2000), version 4.5.3.

To evaluate the performance of the cotton yield prediction model, a linear regression analysis was performed between the estimated cotton boll numbers and the ground truth values. The coefficient of determination (R^2) was used to check how closely the estimations mirrored the actual boll number at the individual plot level. Additionally, to facilitate performance comparison with other yield prediction studies that may use different scales, the normalized root mean squared error (NRMSE) was computed over the range of observed values—maximum cotton boll number minus minimum cotton boll number for the ground truth plots. The residuals were also computed to observe the difference between the ground truth data and the predicted values. Furthermore, to validate the performance of the yield prediction algorithm, the mean absolute percentage error (MAPE) was computed between the predicted number of cotton bolls and the ground truth measurements. These performance indices were computed using the following equations:

$$NRMSE = \frac{\sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}}{y_{max} - y_{min}} \quad (9)$$

$$MAPE(\%) = \frac{1}{N} \times \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (10)$$

where N is the total number of data points used for the linear regression analysis ($N = 45$), y_i is the actual number of cotton bolls for the i^{th} ground truth plot, $\hat{y}_i$ is the number of cotton bolls predicted by the SVM model for the image plot associated to the i^{th} ground truth plot, and $\bar{y}$ is the average number of cotton bolls per plot calculated from the ground truth values of all the 45 ground truth plots.

2.4. Genotype Analysis

The average number of predicted cotton bolls and the standard error (SE) for each genotype and population were calculated to evaluate the statistical accuracy of the yield estimations. To evaluate the effectiveness of the yield estimations, the null hypothesis of equal mean value of yield across all the commercial cultivars and breeding lines was tested using the one-way analysis of variance (ANOVA) at the significance level of 0.05. The statistical computing and graphics software R (R Core Team, 2020), version 4.0.3, was used for this test. After testing the effects due to genotype and its significance, the Fisher's Least Significant Difference (LSD) test was used to judge the likelihood that the observed differences between genotypes and populations comprised non-zero differences in yield performance. The LSD test was performed using the R package *agricolae* (de Mendiburu and Yaseen, 2020), version 1.3-5, to test differences among means of yield for all the genotypes.

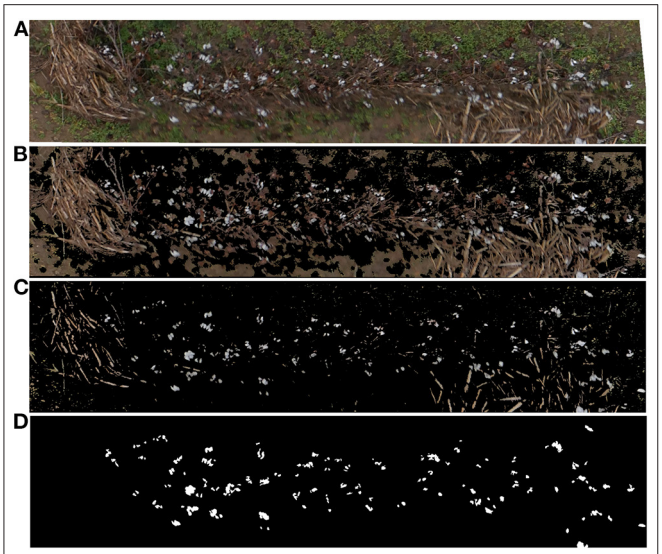


FIGURE 5 | Individual plot image preprocessing results, **(A)** Raw RGB image extracted from the orthomosaic map, **(B)** Preprocessing results for the same plot image after removing vegetation pixels using the $ExGR_{mod}$ index (Equation 1), **(C)** Preprocessing results for the same plot image after removing also bare soil pixels using the $Soil_{Idx}$ index (Equation 2), **(D)** Binary mask used for SVM pixel classifier training. White color identifies cotton pixels. Black color identifies non-cotton pixels.

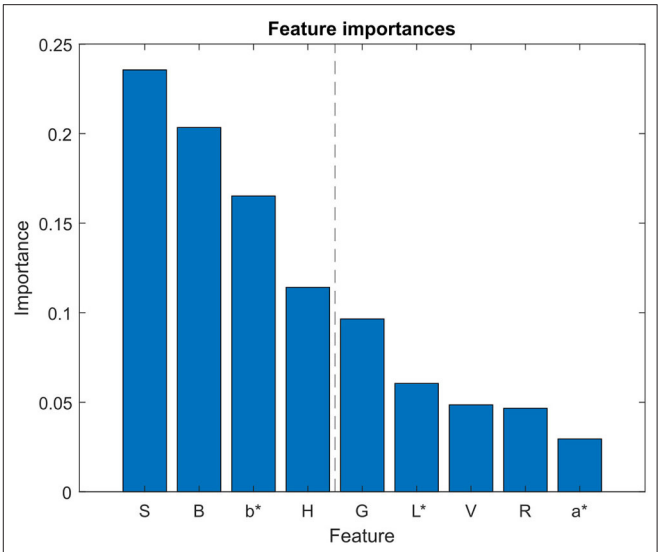


FIGURE 6 | Feature selection analysis. Feature importances obtained using random forest algorithm for feature selection. Blue bars represent the importance of each feature in the classifier model. L*, a*, and b* refer to the CIELAB color space components.

3. RESULTS

3.1. Individual Plot Image Extraction and Preprocessing

A total of 408 valid aerial images were used to generate the orthomosaic map (**Figure 3**). The 45 plots with the associated ground truth data were extracted manually from the orthomosaic map and saved as individual GeoTIF files. **Figure 5** shows one of the extracted plot images. Specifically, this image was used for training the SVM classifier. The raw RGB image (**Figure 5A**) had 523,092 pixels. After removing the vegetation pixels using the $EXGR_{mod}$ index (Equation 1), the new processed image (**Figure 5B**) had 231,447 pixels, which means that the total number of points to analyze was reduced to 44.25%. Finally, after removing the bare soil pixels using our $Soil_{Idx}$ index (Equation 2), the processed image (**Figure 5C**) was down to 34,212 pixels, 6.54% of the raw image.

3.2. SVM Classification Model Development

A single image plot was used to train the SVM model. The image selected for developing the classifier included not only the cotton plants and the cotton bolls, but also other objects typically found in the crop such as old branches and other woody objects from previous crops, weeds, and soil (**Figure 5A**). The result of the annotation process was a binarized image mask in TIF file format (**Figure 5D**).

For feature selection, just the 9 color channels (R, G, B, H, S, V, L*, a*, b*) were analyzed. The RFE algorithm was configured to select the 4 best features, by removing one feature at each iteration using a Random Forest (RF) classifier as the estimator. Results from the RF classifier (**Figure 6**) showed that the 4 most important features were S, B, b*, and H color components. These features were then used to create the training dataset, implying a reduction from the original 11-dimensional feature vector to a 4-dimensional vector. The resulting dataset, which contained 34,212 pixels of the preprocessed training image, was then split into the training subset (27,369 pixels) and the testing subset (6,843 pixels). By reducing the number of features from 11 to 4, the average time needed to classify the pixels of the training image plot was reduced a 13.3%, from 8.42 to 7.3 s.

To analyze qualitatively the results of our cotton pixel classifier, a color code was used to identify image pixels. **Figure 7** shows the inference results of the cotton classifier on the training plot image compared to the annotated mask. As can be seen, most of the cotton pixels are marked with cyan color, which indicates that they were correctly classified by the SVM model, i.e., true positives (**Figures 7B–D**). However, the classifier missed some of the cotton pixels presented in the image. A small portion of real cotton pixels were wrongly classified as non-cotton pixels, comprising false negatives or Type II errors and can be seen as pure blue pixels in **Figures 7B,D**. A smaller portion of non-cotton pixels were misclassified as cotton, comprising false positives or Type I errors and can be identified as pure red pixels in **Figures 7A,D**.

Quantitatively, the classification results of the trained SVM model on the testing subset are summarized in **Figure 8**. The column and row shown in gray indicate the classifier's overall performance. The cell in the bottom right of the plot shows the overall accuracy—correct predictions—of the pixel classifier. The model achieved an accuracy of 88.7%; 22.9% of the 6843

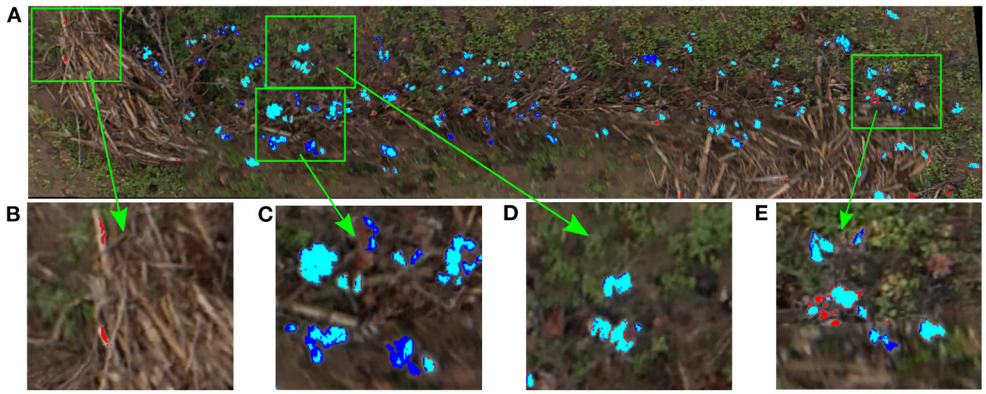


FIGURE 7 | Inference results of the SVM classifier on the training plot image. Blue pixels were missed by the classifier (Type II errors); red pixels were misclassified as cotton by the classifier (Type I errors); cyan were correctly classified cotton pixels, **(A)** Training plot image classification results, **(B)** Non-cotton pixels misclassified as cotton (Type I errors), **(C)** Cotton boll pixels not fully detected (Type II errors), **(D)** Cotton boll pixels fully correctly detected, **(E)** Type I and Type II errors mixed together.

testing pixels were correctly classified as cotton and 65.8% of all testing pixels were correctly classified as non-cotton. Only 11.3% of predictions were wrong; 7.5% of the cotton pixels were incorrectly classified as non-cotton and 3.8% of non-cotton pixels were incorrectly classified as cotton. The column on the far right of the plot shows the percentages of all pixels predicted to belong to each class that were correctly and incorrectly classified. Accounting only for the positive class identification, the upper right cell indicates the precision and the rate of Type I errors of our model. With 1564 out of 1821 cotton pixels being correctly predicted, the SVM classifier achieved a precision of 85.8%, and a Type I error rate of 14.2%. The row at the bottom of the plot shows the percentages of all the pixels that belonging to each class were correctly and incorrectly classified. The bottom left cell indicates the recall and the rate of Type II errors of our classifier. Out of 2079 actual cotton pixels, the model achieved a recall of 75.2% and a Type II error rate of 24.8%. Finally, the image pixel classifier achieved an F1 score of 80.2% at detecting cotton pixels on the testing subset.

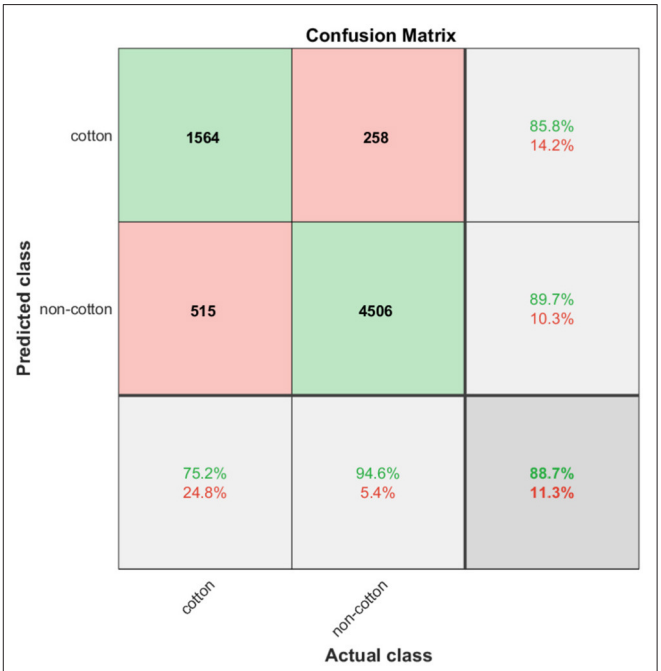


FIGURE 8 | Confusion matrix chart. Green cells show the number of correct classifications by the SVM model. Red cells show the number of classification errors. Gray cells show overall performance of the model.

3.3. Plot-Level Cotton Yield Estimation Model Development

The pixel classifier we developed was then used to extract the cotton pixels from the 45 individual images associated with the digital ground truth plots. It was able to detect subtle color changes, and was robust enough to avoid misclassifying most of the woody elements and the soil (**Figure 9A**). After applying the image post-processing steps and the clustering algorithm, we obtained an estimation of the number of cotton bolls for each image. **Figure 9B** shows the clustering results of a sample image plot extracted from the orthomosaic map. A total of 344 different cotton bolls (pixel clusters) were identified on this particular image plot. Each one of these clusters are identified by a unique color to facilitate the visual analysis.

The number of cotton bolls estimated for the 45 individual image plots analyzed was regressed against its ground truth measurement (**Figure 10**). The estimations of number of cotton bolls at the plot level shows a strong linear relationship (R^2 = 0.932) with the ground truth measurements. This trend is consistent at different numbers of cotton bolls, which indicates that our pixel classifier and clustering algorithm adapted to the changing scenes and was able to segment properly the cotton bolls from both low yielding plots and high yielding plots. The analysis of residuals showed randomly dispersed points around the horizontal axis with no apparent pattern, which indicates that

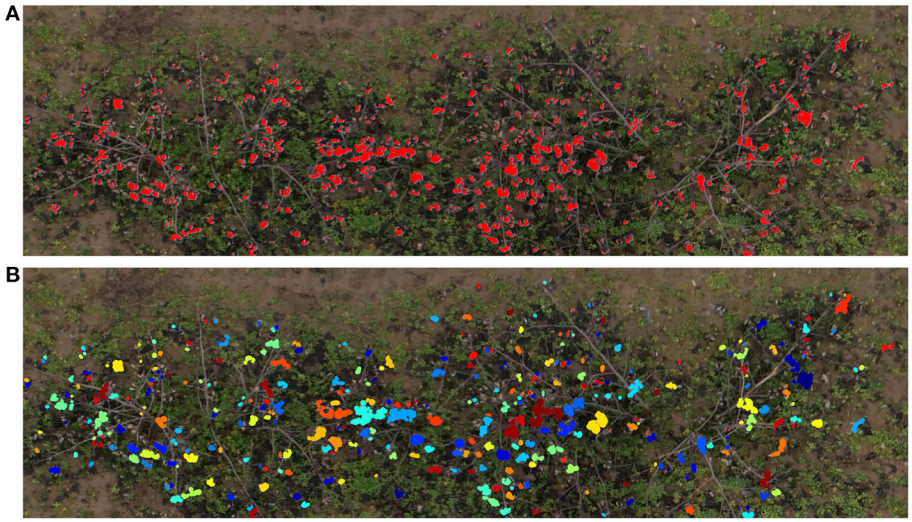


FIGURE 9 | Cotton boll inference results. Cotton pixels segmentation and clustering results for a representative plot (Plot ID 45003), **(A)** SVM classifier inference results. Red pixels represent the cotton pixels detected by the SVM classifier as cotton, **(B)** Cotton boll clustering results. Each cluster is identified by a unique color.

the linear model was a good fit for the input data. Only one data point does not follow the regular distribution (red circled point on **Figure 10**) and given its value is more than three standard deviations from the mean, it was identified as an outlier. Our model achieved a MAPE of 13.672% at detecting cotton bolls and a normalized RMSE value of 0.066 over the range of observed cotton bolls.

3.4. Web API Deployment

A web-based API (web app) was developed to integrate the developed pipeline into a more usable interface with the aim of improving automation and usability (**Figure 11**). The web app consisted of three basic functions to process each input plot image: the preprocessing steps (vegetation and soil pixels removal), the SVM classifier deployment (creation of features and SVM pixel classification), and the final cotton boll number estimation (morphological image processing operations and connected components labeling algorithm). The SVM classifier and the clustering algorithm were deployed using Flask as the core of the API, in a dockerized environment. A Docker container image of the web app is available on Docker Hub in the repository https://hub.docker.com/r/javirodsan/yieldestimation. Additionally, we will provide the code and some sample images for testing at https://github.com/Javi-RS/Cotton_Yield_Estimation.

3.5. Genotype Analysis Results

The mean number of predicted cotton bolls and the SE for each genotype are summarized in **Table 2**. Results show that the cotton yield estimations produced by the proposed method has relatively low SEs for each cultivar and breeding line, which indicates that the means of the yield for the different genotypes are centered around the population mean, and hence, the sampled plots are representative of the population.

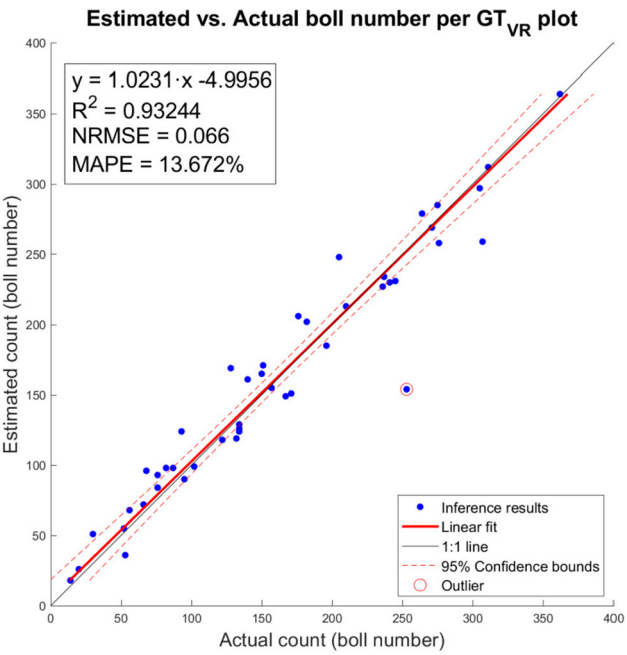


FIGURE 10 | Estimated vs. actual number of cotton bolls per ground truth plot (GT_{VR}). Blue dots represent inference results of our model. Red line represents the linear fit. Red dotted lines represent the 95% confidence interval for the fit. One outlier is marked with a circle surrounding the data point.

The ANOVA test identified significant differences between the means of estimated yield for cotton genotypes, with $F_{(15,\ 472)} = 1.874$ and $p < 0.05$. Thus, the null hypothesis of equal mean value of yield across all the genotypes can be rejected, which suggests that our methodology was effective in identifying differences in

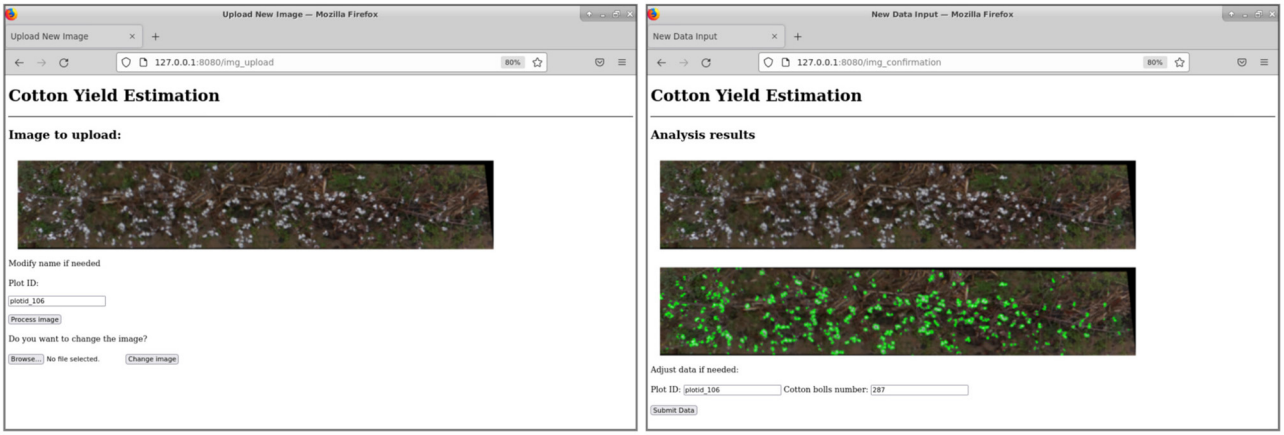


FIGURE 11 | Web API for automatic image analysis.

TABLE 2 | Statistical analysis summary of predicted yield for commercial cultivars and breeding lines.

Genotype	Sample size	Mean (boll number)	SE	Groups
UA48	10	237.300	26.755	a
Acala1517-98	10	211.500	25.937	ab
GA230	10	199.100	21.969	abc
DeltaPine393	10	198.300	23.137	abc
L†	16	179.625	15.264	abcd
N†	20	178.800	32.627	bcd
FM832	10	176.800	19.960	bcd
Q†	40	172.525	11.224	bcd
S†	184	166.576	5.380	bcd
O†	22	163.591	18.585	bcd
M†	8	158.875	32.627	bcd
K†	44	158.364	10.682	cd
TAM94L25	10	156.300	18.049	cd
R†	44	151.614	10.072	cd
J†	18	143.889	14.703	cd
P†	32	137.594	11.275	d

Data are sorted from higher to lower yield by genotype. Genotypes are grouped according to the probability of means differences and alpha level (0.05). Cultivars and breeding lines with the same letter are not significantly different. †Indicates breeding line populations, comprised of samples of progeny from crosses between different mutant lines described in Patel et al. (2014).

yield. The Fisher's LSD identified statistical differences between the estimation of average yield for the cultivars, with UA48 and Acala1517-98 being significantly higher than TAM94L25, while FM832 being significantly lower than UA48. Regarding the mutant-derived populations, L and N populations had the highest yielding, and P had the lowest yielding, with other groups in-between. While the LSD test was able to identify significant differences of means of yield among the commercial cultivars, individual breeding lines had only two replications which provided insufficient evidence for definitive ranking.

4. DISCUSSION

Estimating cotton yield before harvesting would assist breeders to identify highly productive genotypes without incurring the time and cost of actually harvesting the field. Our study demonstrated that the number of cotton bolls present on each individual plot of a field can be estimated accurately by using RGB images captured from a drone flight at a low altitude. This approach can be used to quickly estimate yield at the plot level and would allow cotton breeders analyze large variety trials efficiently, especially with higher levels of replication as were used for the cultivars. However, additional data would be needed to confirm its usability on experimental breeding lines.

4.1. Comparison With Other Studies

As opposed to previous methods, our method used a supervised machine learning model to classify the pixels in the image instead of using traditional global thresholding techniques. Approaches for cotton yield estimation based on traditional image processing techniques (Huang et al., 2016; Yeom et al., 2018; Dodge, 2019) usually assume that cotton has a distinctive spectral response that enables the easy discrimination of cotton bolls from the rest of the elements of the crop just by using a threshold value in one or more of the RGB channels. However, in a real-case scenario the illumination conditions can change considerably during data collection, and often the range of image intensities of the color channels for the cotton bolls are similar to other crop elements. Although some studies have applied adaptive threshold techniques or have included prior preprocessing steps to minimize the limitations of global thresholding techniques (Maja et al., 2016), these approaches are not flexible enough to cope with the variability of reflectance across a field, and hence their accuracy is limited. Our method uses an SVM model based on 4 image channels to segment cotton-related pixels. Machine learning techniques are more flexible than traditional image processing methods at finding patterns on data with non-obvious relationships. A recent study (Ashapure et al., 2020) has already investigated the use of machine learning techniques to

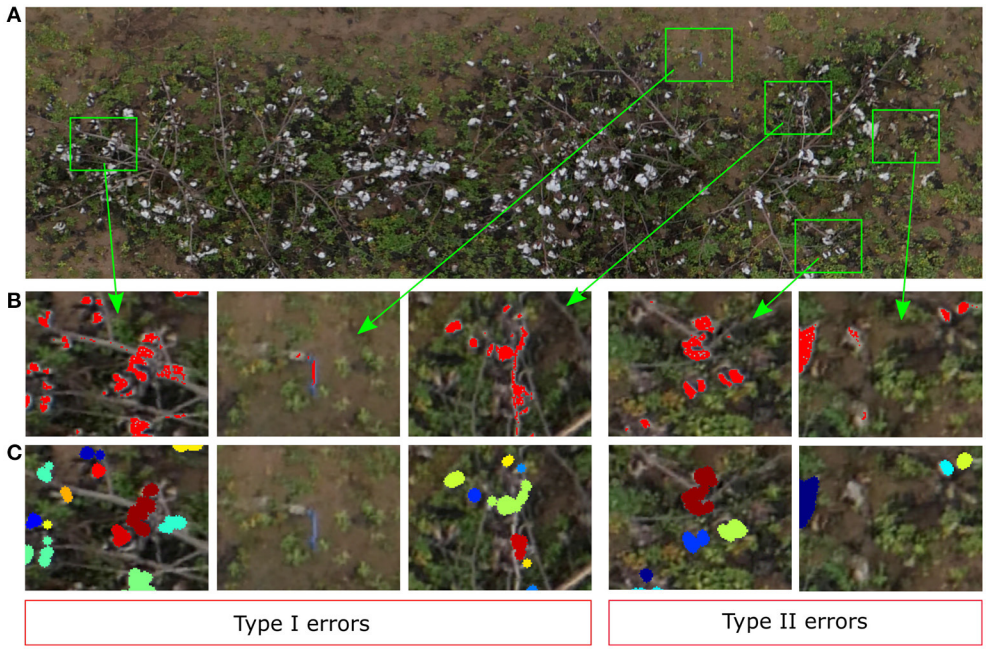


FIGURE 12 | Qualitative analysis of the yield estimator. **(A)** Original image plot (Plot ID 45003), **(B)** Zoomed-in view to Type I and Type II errors at segmenting cotton pixels. Red represents the pixels segmented by the SVM classifier as cotton pixels, **(C)** Zoomed-in view to the same zones as **(B)** to visualize the clustering results. Different colors identify different clusters (cotton bolls).

estimate cotton yield. However, the focus of their study was to find the relationship between cotton yield and the parameters of the crop during the course of the season, not near harvest. Therefore, the potential of this approach for later growth stages might be limited because it includes crop features related to the canopy status. Our approach was developed to be applied after crop defoliation (which commonly precedes cotton harvest) to reduce the effects of occlusions by leaves and maximize cotton boll visibility. Although this limits the applicability of our method to the pre-harvesting time frame, this is the period during which breeders evaluate the overall performance of new breeding lines, so it can be considered one of the key stages in the selection trials.

4.2. Type I and II Errors Analysis

Our SVM classifier was able to segment the cotton pixels accurately from the input images, showing a promising overall performance for the training image (**Figure 7**). As we have commented in section 3.2, the classifier made some Type I errors (false positives), and Type II errors (false negatives). In our context we tried to minimize the Type I errors, i.e., the number of background pixels wrongly classified as cotton pixels. We aimed to detect all the cotton pixels in the image, but we did not want to overestimate them. Usually, the number of cotton boll pixels in a plot image is much lower than the number of pixels from other parts of the plants and background. Hence, the chances of misclassifying non-cotton pixels are higher. **Figure 12** shows the performance of the SVM classifier model on an unseen plot image. Type I errors were mainly caused by elements of the scene with a spectral response similar to the cotton bolls. As shown in **Figure 12B**, some branches and other foreign objects in the field were highly reflective and they were misclassified as cotton pixels.

Type II errors were caused primarily by dark cotton pixels in the image that were not properly detected by the classifier. Bolls from the lower parts of the canopy were less reflective than those from the top parts with more light. Therefore, the image pixels were usually darker on these parts, and the SVM classifier was not able to detect completely all the cotton pixels of some cotton bolls (**Figure 12B**). These zones are usually small, and the image post-processing operations tend to remove them before clustering (**Figure 12C**). Even though there were some obvious classification errors, most of the cotton bolls detected by our algorithms were true cotton bolls. These errors do not necessarily have a substantial effect on our estimations of yield because it was expected that not all the cotton bolls can be seen from downward images.

The morphological operations enhanced the appearance of the cotton bolls in the binarized images and reduced the influence of the SVM Type I errors on the cotton boll estimation. The small elliptic structuring element used for the erosion operation contributed to removing isolate pixels and small pixel clusters associated to branches and other wrongly classified elements, and hence the pixel clustering step performed relatively well at isolating cotton bolls and correcting Type I errors (**Figure 12C**). A larger element for the dilation operation aided to extend the boundaries of the rest of cotton pixels to eliminate gaps between close pixels and consolidate cotton bolls (**Figure 12C**).

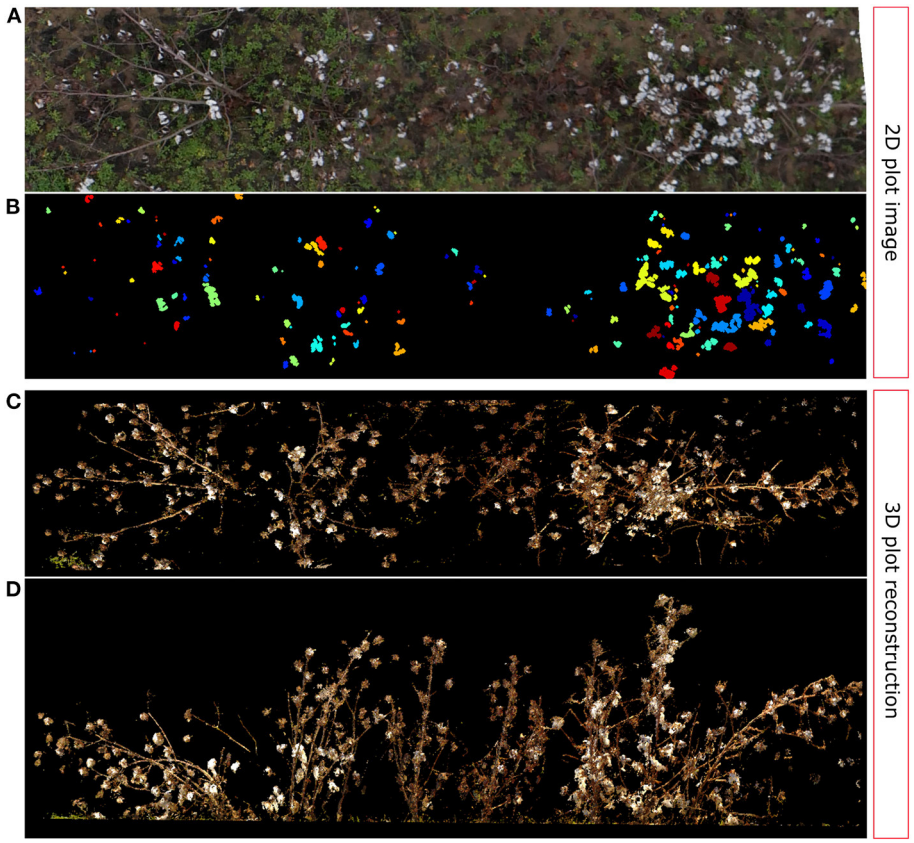


FIGURE 13 | Outlier analysis: 2D vs. 3D comparison. **(A)** Original image plot (Plot ID 41028), **(B)** Cotton boll clustering results; each cluster is identified by a unique color, **(C)** TLS reconstructed 3D point cloud: Top view, **(D)** TLS reconstructed 3D point cloud: Frontal view.

4.3. Limitations

We used a digital RGB camera to collect the aerial images, which internally processes the data from the image sensor and performs the JPEG compression to save the image in a removable storage media. These images are stored without a previous radiometric correction. We tried to minimize the effect of the illumination on the data collection day by using an automatic color balance compensation. Although the AWB compensation was adequate for our data analysis, it might not be a universal solution for all the possible illumination conditions in the field. Therefore, the use of our model directly to images collected at different times may be limited if the atmosphere and solar radiation greatly differ from those on our data collection day. However, since the method we proposed is relatively fast and easy to use, retraining the SVM model with new data from the specific collection day can be feasible.

Additionally, as we noted in section 3.3, one of the points in the data set was identified as an outlier during the linear regression analysis (red circled point in **Figure 10**). By further analyzing this particular data point and the associated plot (**Figure 13**), we can determine that these kind of errors are caused by one of the main limitations of 2D image analysis: lose of depth information. The SVM classifier detected the cotton pixels in the orthomosaic image fairly well. Moreover, the clustering algorithm was able to find and segment properly some of the cotton bolls (**Figure 13B**). However, the high density of cotton bolls in this plot made the cotton pixels to appear close together on the 2D aerial image (**Figure 13A**), which prevented our clustering algorithm from segmenting all the cotton bolls properly. Therefore, the estimation of number of cotton bolls for this plot was not accurate—only 154 out of 253 cotton bolls were detected. If we compare this image with its 3D point cloud counterpart, we can see that a substantial number of the cotton bolls were located in almost vertical branches (**Figure 13D**), which made the lower cotton bolls to be heavily occluded by the rest when looking from the top (**Figure 13C**). In a 2D image all the pixels are contained in the same plane. This lack of depth information impeded our algorithm to identify cotton pixels at different height levels and led to underestimating the number of cotton bolls.

4.4. Future Work

Most of the image processing can be performed automatically without any supervision. However, the extraction of plot images from the orthomosaic map was carried out manually. A future research direction could investigate the feasibility of automatically separating the plots from the orthomosaic map using only geographic information. This would improve

the efficiency of this methodology and would contribute to improving the throughput for breeding purposes. Additional research could also be performed to improve cotton pixel classification, including the use of more advanced deep learning methods such convolutional neural networks (CNN) that could discriminate cotton pixels from the rest of elements by automatically integrating spatial and morphological information as additional features without needing to design them manually. Even though the methodology presented here was developed and tested using individual plot images, it could be modified easily to perform yield estimations on production fields because of its ease of use. Some minor modifications on the processing pipeline would make it suitable for production fields. Instead of extracting entire plots, the orthomosaic map could be divided into several sections or cells using grids with fixed dimensions. Then, our method could be used to estimate the yield from each cell and then aggregate the estimations to have an estimation of the total production of the entire field. However, further studies will need to be carried out to validate this approach on different crop densities. Additionally, our model was able to detect cotton bolls consistently from almost all the plots of the field. However, we didn't correct the images atmospherically, which may affect the direct application of our model to images collected under different conditions. We will investigate how to further improve our pipeline by including a preprocessing step to correct radiometrically the images and make our method agnostic to illumination lighting conditions and imaging sensors. Finally, we were able to use digital ground truth techniques based on 3D information to demonstrate some of the limitations of 2D approaches for estimating yield from aerial imagery. We will investigate the feasibility of using 3D crop analysis to overcome these limitations and its viability to estimate other crop traits.

5. CONCLUSIONS

This study presents a cost-effective approach for estimating cotton yield production from images collected using a drone and a conventional RGB camera. A supervised machine learning classifier based on an SVM model was trained using only a single plot image. Since this approach requires the annotation of only one RGB image, it reduces the complexity and time needed for model deployment. The classifier demonstrated to be robust to changing scenes and discriminated accurately the cotton pixels in individual plot images with different number of cotton bolls. Consequently, reliable cotton boll counting was obtained. In addition, the methodology was found to be effective in identifying the differences in yield among different commercial cultivars and breeding lines. Overall, the proposed method can help improve the efficiency of decision making for breeding programs and optimize the use of resources by speeding up the analysis of entire field trials. Future work will focus on automating the extraction of plot images, as well as on the application of 3D-based approaches and more advanced machine learning methods to improve cotton boll detection.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JR-S contributed to the conceptualization of the idea, methodology, investigation, software development, validation, formal analysis, data curation, writing—original draft preparation, visualization, and project administration. CL contributed to the conceptualization, methodology, provision of resources, supervision, project administration, and funding acquisition. AP contributed to the methodology, provision of resources, and funding acquisition. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was funded by the National Science Foundation Growing Convergence Research (Award No. 1934481) and the Georgia Cotton Commission.

ACKNOWLEDGMENTS

The authors gratefully thank Rui Xu for his assistance in collecting aerial images; Tsunghan Han, Shangpeng Sun, and Jon S. Robertson for helping in the manual harvesting of the cotton bolls for ground truthing; Man Wah Aw for his assistance in image annotation; and Ellen Skelton for maintaining the crop. The authors also thank Dr. Kyle Johnsen and Anton Franzluebbers from the Virtual Experience Laboratory at the University of Georgia for providing us with the VR annotation toll for digital ground truth counting.

REFERENCES

Ashapure, A., Jung, J., Chang, A., Oh, S., Yeom, J., Maeda, M., et al. (2020). Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. *ISPRS J. Photogrammetry Remote Sens.* 169, 180–194. doi: 10.1016/j.isprsjprs.2020.09.015

Barbedo, J. G. A. (2019). A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses. *Drones* 3. doi: 10.3390/drones3020040

Bourland, F., and Myers, G. O. (2015). "Conventional cotton breeding," in *Cotton*, eds D. D. Fang and R. G. Percy (Madison, WI: John Wiley & Sons, Ltd.), 205–228. doi: 10.2134/agronmonogr57.2013.0025

Bowman, D., Bourland, F., Myers, G., Wallace, T., and Caldwell, D. (2004). Visual selection for yield in cotton breeding programs. *J. Cotton Sci.* 8, 62–68.

Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal of Software Tools. Available online at: https://github.com/opencv/opencv/wiki/CiteOpenCV

Chu, T., Chen, R., Landivar, J., Maeda, M., Yang, C., and Starek, M. (2016). Cotton growth modeling and assessment using unmanned aircraft system visual-band imagery. *J. Appl. Remote Sens.* 10, 1–17. doi: 10.1117/1.JRS.10.036018

Cortes, C., and Vapnik, V. (1995). Support vector networks. *Mach. Learn*. 20, 273–297. doi: 10.1007/BF00994018

de Mendiburu, F., and Yaseen, M. (2020). *agricolae: Statistical Procedures for Agricultural Research. R package version 1.4.0*. Available online at: https://cran.r-project.org/web/packages/agricolae/index.html

Dodge, W. (2019). *Image based yield estimation in cotton using UAS* (Ph.D. dissertation). Texas Tech University, Lubbock, TX, United States.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., and Vapnik, V. (1996). "Support vector regression machines," in *NIPS*, eds M. Mozer, M. I. Jordan, and T. Petsche (Cambridge, MA: MIT Press), 155–161.

Feng, A., Zhou, J., Vories, E. D., Sudduth, K., and Zhang, M. (2020). Yield estimation in cotton using UAV-based multi-sensor imagery. *Biosyst. Eng*. 193, 101–114. doi: 10.1016/j.biosystemseng.2020.02.014

Garcia-Ruiz, F., Sankaran, S., Maja, J. M., Lee, W. S., Rasmussen, J., and Ehsani, R. (2013). Comparison of two aerial imaging platforms for identification of huanglongbing-infected citrus trees. *Comput. Electron. Agric*. 91, 106–115. doi: 10.1016/j.compag.2012.12.002

Hdioud, B., Tirari, M. E. H., Thami, R. O. H., and Faizi, R. (2018). Detecting and shadows in the HSV color space using dynamic thresholds. *Bull. Electric. Eng. Inform*. 7, 70–79. doi: 10.11591/eei.v7i1.893

Huang, Y., Howard, J. B., Ruixiu, S., Steven, J. T., Tomonari, F., and, M., et al. (2016). Cotton yield estimation using very high-resolution digital images acquired with a low-cost small unmanned aerial vehicle. *Trans. ASABE* 59, 1563–1574. doi: 10.13031/trans.59.11831

Huang, Y., Sui, R., Thomson, S., and Fisher, D. K. (2013). Estimation of cotton yield with varied irrigation and nitrogen treatments using aerial multispectral imagery. *International J. Agric. Biol. Eng*. 6, 37–41. doi: 10.3965/j.ijabe.20130602.005

Jung, J., Maeda, M., Chang, A., Landivar, J., Yeom, J., and McGinty, J. (2018). Unmanned aerial system assisted framework for the selection of high yielding cotton genotypes. *Comput. Electron. Agric*. 152, 74–81. doi: 10.1016/j.compag.2018.06.051

Liu, S., and Whitty, M. A. (2015). Automatic grape bunch detection in vineyards with an svm classifier. *J. Appl. Log*. 13, 643–653. doi: 10.1016/j.jal.2015.06.001

Maja, J. M. J., Campbell, T., Camargo Neto, J., and Astillo, P. (2016). "Predicting cotton yield of small field plots in a cotton breeding program using UAV imagery data," in *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping*, eds J. Valasek and J. A. Thomasson (Baltimore, MD), 98660C. doi: 10.1117/12.2228929

Meyer, G., and Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric*. 63, 282–293. doi: 10.1016/j.compag.2008.03.009

Patel, J. D., Wright, R. J., Auld, D., Chandnani, R., Goff, V. H., Ingles, J., et al. (2014). Alleles conferring improved fiber quality from EMS mutagenesis of elite cotton genotypes. *Theoret. Appl. Genet*. 127, 821–830. doi: 10.1007/s00122-013-2259-6

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

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Raza, S.-A., Prince, G., Clarkson, J. P., and Rajpoot, N. M. (2015). Automatic detection of diseased tomato plants using thermal and stereo visible light images. *PLoS ONE* 10, e0123262. doi: 10.1371/journal.pone.0123262

Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., and Plümer, L. (2010). Original paper: early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Comput. Electron. Agric*. 74, 91–99. doi: 10.1016/j.compag.2010.06.009

Song, Y., Glasbey, C. A., Horgan, G. W., Polder, G., Dieleman, J. A., and van der Heijden, G. (2014). Automatic fruit recognition and counting from multiple images. *Biosyst. Eng*. 118, 203–215. doi: 10.1016/j.biosystemseng.2013.12.008

Townsend, T., and Sette, J. (2016). "Natural fibres and the world economy," in *Natural Fibres: Advances in Science and Technology Towards Industrial Applications*, eds R. Fangueiro and S. Rana (Dordrecht: Springer), 381–390. doi: 10.1007/978-94-017-7515-1_30

Vapnik, V., Golowich, S. E., and Smola, A. J. (1997). "Support vector method for function approximation, regression estimation and signal processing," in *Advances in Neural Information Processing Systems 9 – Proceedings of the 1996 Neural Information Processing Systems Conference (NIPS 1996)*, eds M. Mozer, M. I. Jordan, and T. Petsche (Dever, CO; Cambridge, MA: MIT Press), 281–287.

Xia, L., Zhang, R., Chen, L., Huang, Y., Xu, G., Wen, Y., et al. (2019). Monitor cotton budding using SVM and UAV images. *Appl. Sci*. 9. doi: 10.3390/app9204312

Xu, R., Li, C., and Paterson, A. H. (2019). Multispectral imaging and unmanned aerial systems for cotton plant phenotyping. *PLoS ONE* 14, e0205083. doi: 10.1371/journal.pone.0205083

Yeom, J., Jung, J., Chang, A., Maeda, M., and Landivar, J. (2018). Automated open cotton boll detection for yield estimation using unmanned aircraft vehicle (UAV) data. *Remote Sens*. 10. doi: 10.3390/rs10121895

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fpls.2022.851245

Rapid Detection of Wheat Ears in Orthophotos From Unmanned Aerial Vehicles in Fields Based on YOLOX

*Yao Zhaosheng[1,2], Liu Tao[1,2], Yang Tianle[1,2], Ju Chengxin[1,2] and Sun Chengming[1,2]**

[1]*Jiangsu Key Laboratory of Crop Genetics and Physiology/Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College of Yangzhou University, Yangzhou, China,* [2]*Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou, China*

Wheat ears in unmanned aerial vehicles (UAV) orthophotos are characterized by occlusion, small targets, dense distribution, and complex backgrounds. Rapid identification of wheat ears in UAV orthophotos in a field environment is critical for wheat yield prediction. Three improvements were achieved based on YOLOX-m: mosaic optimized, using BiFPN structure, and attention mechanism, then ablation experiments were performed to verify the effect of each improvement. Three scene datasets were established: images were acquired during three different growing periods, at three planting densities, and under three scenarios of UAV flight heights. In ablation experiments, three improvements had increased recognition accuracies on the experimental dataset. Compared the accuracy of the standard model with our improved model on three scene datasets. Our improved model during three different periods, at three planting densities, and under three scenarios of the UAV flight height, obtaining 88.03%, 87.59%, and 87.93% accuracies, which were, respectively, 2.54%, 1.89%, and 2.15% better than the original model. The results of this study showed that the improved YOLOX-m model can achieve UAV orthophoto wheat recognition under different practical scenarios in large fields, and that the best combination were obtained images from the wheat milk stage, low planting density, and low flight altitude.

Keywords: small target, spike, YOLOX, UAV, Orthophoto, BiFPN

OPEN ACCESS

Edited by:
Daobilige Su,
China Agricultural University, China

Reviewed by:
Guo Leifeng,
Agricultural Information Institute (CAAS), China
Zhenwang Qin,
Agricultural Information Institute (CAAS), China

****Correspondence:***
Sun Chengming
dx120200098@stu.yzu.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *09 January 2022*
Accepted: *05 April 2022*
Published: *27 April 2022*

Citation:
Zhaosheng Y, Tao L, Tianle Y, Chengxin J and Chengming S (2022) Rapid Detection of Wheat Ears in Orthophotos From Unmanned Aerial Vehicles in Fields Based on YOLOX. Front. Plant Sci. 13:851245. doi: 10.3389/fpls.2022.851245

INTRODUCTION

Wheat yield is calculated from the number of spikes per unit area, the number of grains per spike, and the weight of grains. In agricultural production, especially in wheat cultivation and breeding, determination of the number of spikes per unit area still relies on manual work, which introduces human error during prolonged, intensive work. Therefore, a fast, accurate method for counting wheat spikes in a large field environment is essential.

Compared with the time-consuming and laborious manual counting, modern information tools such as machine-learning methods, image analysis techniques, and artificial intelligence technologies can significantly improve the efficiency of wheat spike counting (Xu et al., 2020). Research on wheat-ear recognition has developed considerably in the last decade and can be broadly classified into traditional image processing methods, deep-learning network methods, and other methods. Using first-order and high-order methods, Frdric et al. (2012) attempted

to extract wheat-ear texture features and used the K-means algorithm for ear segmentation and wheat-ear counting by setting thresholds. Cointault and Gouton (2007) combined color features based on texture features for wheat-ear segmentation and used the skeleton method to solve the problem of overlapping wheat ears during counting to improve the accuracy of wheat-ear recognition. Tao et al. (2014) used image processing techniques to complete the segmentation of wheat ears in a large field, and added the corner-point detection method to achieve automatic counting of wheat ears. Fernandez-Gallego et al. (2020) utilized a drone to collect images of wheat fields and compared the effectiveness of eight machine learning methods on wheat-ear recognition. Their results showed that the random forest method had the highest accuracy.

Owing to broad applicability and high accuracy, deep-learning methods have become a new means to address the challenges encountered in wheat counting. Sadeghi-Tehran et al. (2019) constructed wheat feature models and fed the models into convolutional neural networks to achieve semantic segmentation and automatic counting of wheat. In addition, TasselNetv2 (Xiong et al., 2019), mobileNetV2 (Khaki et al., 2021), YOLOV4 (Yang et al., 2021), EfficientDet (Wang et al., 2021), LPNet (Misra et al., 2020), and other deep-learning networks have shown advantages in wheat counting.

The challenges to effective wheat detection have promoted the rapid development of machine learning algorithms in wheat-head detection. Many wheat datasets have emerged, among which the most popular one is the Global Wheat Head Detection (GWHD) dataset (David et al., 2021). These developments have played a significant role in advancing wheat-detection algorithms. The equipment used to acquire wheat pictures in this dataset was a digital camera, shooting from a height between 1.8 and 3 m above the ground and a ground sampling distance (GSD) of 0.10–0.62 mm/px.

Relative size is mainly used to define small targets in target detection. The relative size is defined according to the Society of Photo-Optical Instrumentation Engineers (SPIE), and a small target is defined as a target area of fewer than 80 pixels in a 256 × 256 image, i.e., less than 0.12% of 256 × 256 is a small target (Yang et al., 2016). Most current research vehicles for wheat-ear detection are far beyond the small target range: high-definition images are obtained by high-definition cameras hand-held or mounted on a shelf and photographed at a closer distance, which makes the features of wheat ears clear and easy to be extracted to more features by deep networks, and the recognition accuracy can reach 98% (Zhou et al., 2018; Xiong et al., 2019; Li et al., 2021). However, this method does not apply to actual production practices and does not achieve field-wide or larger-scale wheat-ear detection. After wheat heading completed, clear images need to be captured using unmanned aerial vehicles (UAV) and using orthophoto stitching technology, then detect wheat ears.

The ground sampling distance of wheat-ear images obtained by UAV is greatly affected by the flight height of the UAV. Taking DJI Inspire2 as an example, the experiment showed that after wheat head was presented, visible, and fully emerged. The UAV flight height was below 10 m, the wind generated by the propeller blew the ears of wheat about, making them shake and thus affecting the clarity of shooting and multi-photograph synthesis of orthophoto images. Flying the UAV too high made it impossible to extract wheat-ear features by a deep-learning network, yielding poor results. Therefore, a better method was needed for large-scale UAV orthophoto detection of small wheat-ear scenes with low GSD, high density, and small targets.

YOLOX is a series of YOLO improvement algorithms introduced by Kuang-Shi Technology (MEGVII) in 2021 (Ge et al., 2021). YOLOX provides the following improvements over YOLOv3: (1) decoupled head (by decoupling the prediction branches, the convergence speed improves, as does AP by 4.2%, over the non-decoupled end-to-end method); (2) data augmentation (using Mosaic and Mixup and turning off data augmentation for the last 15 epochs to prevent excessive data augmentation); and (3) anchor improvement (using Anchor-free, improving multi-positives and SimOTA, reducing training time, and improving prediction accuracy). This is a good new model. Currently it is not used much in articles. Panboonyuen et al. (2021) utilizing pre-training Vision Transformer (ViT) as a backbone, apply Feature Pyramid Network (FPN) decoder detection of Road Assets, It significantly outperforms other state-of-the-art (SOTA) detectors. Zhang et al. (2021) used the YOLOX algorithm to detect vehicle targets in UAV images, and through a self-made dataset, the detection results surpassed traditional algorithms. At present, there is no article using YOLOX to detect wheat ears.

In this paper, we propose a method to obtain large-scale orthophotos of wheat fields using UAVs with telephoto lenses. We validate the algorithm for wheat-spike detection performance phenotypes under three periods, three densities, and three flight height scenarios using the improved You Only Look Once (YOLOX) deep-learning network algorithm. The improved YOLOX-m model achieves a good effect on low-resolution images. Realize the identification of dense small target wheat ears in large size (1,280pixels × 1,280pixels) images, it's favorable to the identification of wheat ears in large field orthophotos. And the best UAV orthophoto recognition is obtained from the wheat milky stage, low planting density, and low flight altitude. The improved YOLOX model exhibits the higher classification accuracy and the different scene adaptation capability.

MATERIALS AND METHODS

Experimental Designs

This study was conducted at the Fengling Experimental Base of Yangzhou University in Yangzhou City, Jiangsu Province, China (32° 30′ 7″, 119° 13′ 54″) using a 75 m × 25 m field size, with each plot measuring approximately 20 m^2. Wheat plant conformation and spike morphology are influenced by variety. Yangmai 23, which has a large planting area, was selected as the experimental variety, with three densities: 1.2×10^6/ha (D1), 1.8×10^6/ha (D2), and 2.7×10^6/ha (D3), and replicated three times (**Figure 1**). The sowing date was October 11, 2020, the planting method was mechanical strip sowing, fertilization was consistent, and other cultivation measures were consistent with local customs.

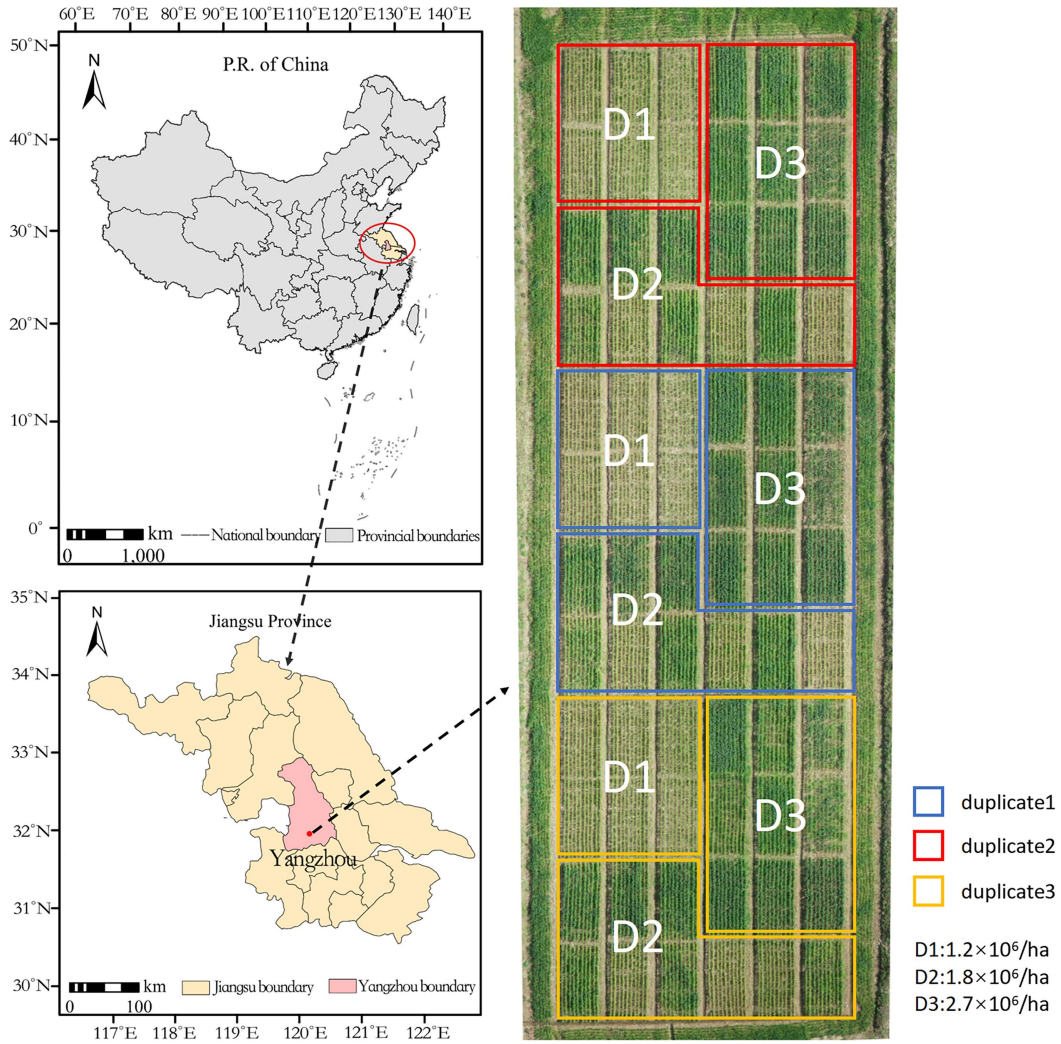


FIGURE 1 | Distribution of test sites and test fields.

Image Acquisition

This study used a DJI insprire2 (inspire2, from SZ DJI Technology Co. Ltd., Shenzhen, China) multi-rotor UAV equipped with a Zenmuse X5S (SZ DJI Technology Co. Ltd., Shenzhen, China) camera, an Olympus M. Zuiko 45 mm/1.8 (Olympus Co. Ltd., Tokyo, Japan) lens (**Figure 2A**), and DJI GS Pro (SZ DJI Technology Co. Ltd., Shenzhen, China) ground station software to conduct one UAV mission at the flowering (P1), milking (P2), and maturity (P3) stages of wheat. The flight parameters were set at 78% heading overlap, 80% bypass overlap, the flight height at 20 m (H1), 25 m (H2), and 30 m (H3). The orthophoto reconstruction of the acquired flight data was performed using DJI Terra (SZ DJI Technology Co. Ltd., Shenzhen, China), and orthophotos were exported for the next step (**Figure 2B**).

Scene Dataset Production

According to the experimental design, three wheat scenes were divided into different periods of wheat scenes (P series), different densities of wheat scenes (D series), and different resolutions of wheat scenes (H series):

P series: D2 image region orthophotos of three periods from P1 to P3 were selected as the image source,

D series: P2 period image orthophotos under three density treatments from D1 to D3 are selected as the image source,

H series: P2 period image orthophotos of three flight heights from H1 to H3 are selected as the image source.

To improve the efficiency of image cropping, a software "Crop Assistant" (**Figure 2C**) was developed to quickly crop the image to a specified size. The mouse is used to let the cross-auxiliary line move to the image area to be cropped, followed by clicking at the center point. The image size of 200 × 200 pixels can be intercepted with the cross as the center point, and the image is automatically named and saved to a preselected folder according to the rules (**Figure 2D**). The user then manually labels the cropped images with wheat ears and generates the corresponding xml file, which contains information

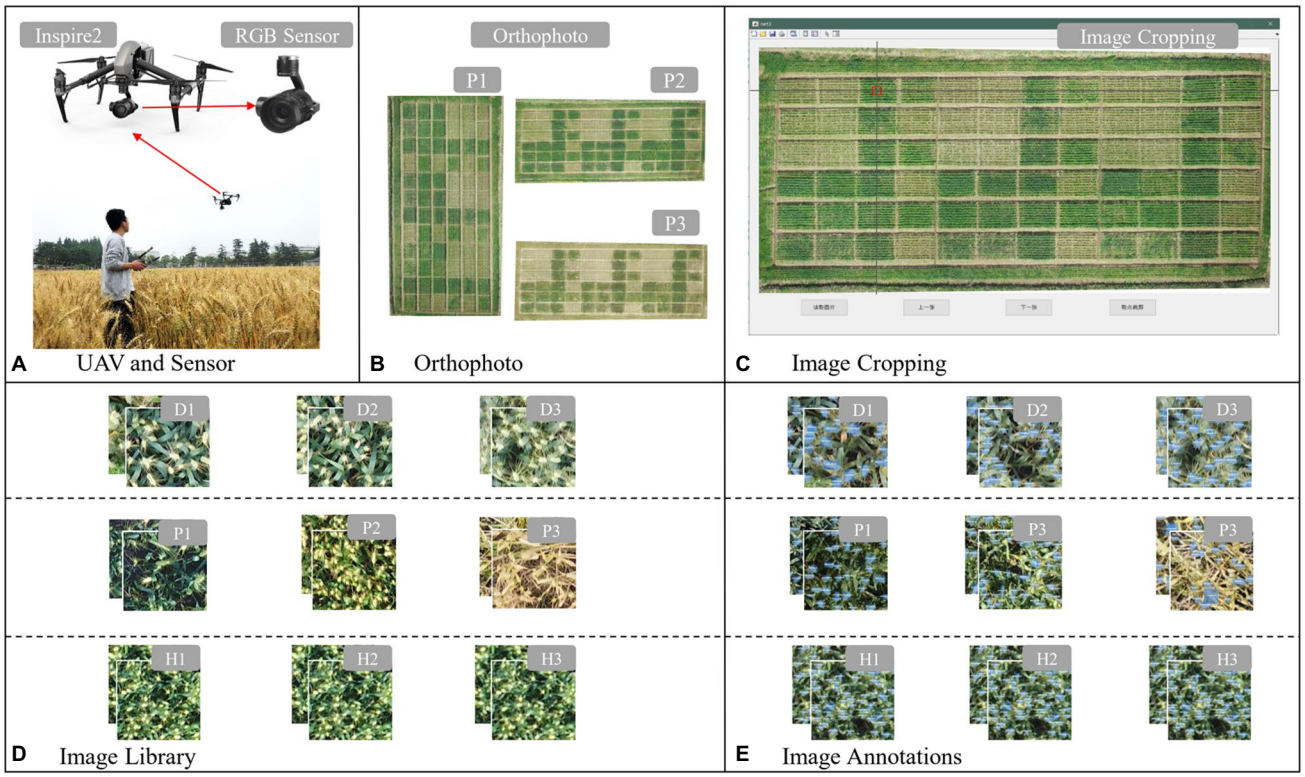


FIGURE 2 | Image acquisition and dataset creation methods: **(A)** unmanned aerial vehicle (UAV) and sensor, **(B)** orthophoto, **(C)** image cropping, **(D)** image library, and **(E)** image annotations.

TABLE 1 | Dataset base information.

Dataset	P Series			D Series			H Series			Total
	P1	P2	P3	D1	D2	D3	H1	H2	H3	
N_{img}	370	395	390	368	355	350	302	327	314	3,171
N_{lab}	8,547	9,717	9,321	7,544	8,591	10,640	7,399	9,810	11,304	82,873

N_{img}, total number of images and N_{lab}, total number of label boxes.

such as image size, label name, and target location (**Figure 2E**). The number of statistical images (N_{img}) and the total number of labeled boxes (N_{lab}) are shown in **Table 1**. There were 3,171 images labeled in three series of nine-scene datasets, with a total number of wheat ears of 82,873. The amount of data met the number of datasets required by a deep-learning network. Each dataset is randomly divided into a training set, a validation set, and a test set in a 7:2:1 ratio for model training.

YOLOX-m MODEL IMPROVEMENT AND EVALUATION METRICS

YOLOX has a good recognition effect on target in the existing model, but there is still room for improvement. Sun et al. (2021) compared the performance of YOLOX and Deformable DETR (Deformable transformers for end-to-end object detection) models in the identification of bok choy seedlings from Weeds. The results showed that YOLOX was the optimal model, and got better average precision and identification speed. In order to better apply YOLOX to wheat ear detection, we have made three improvements: (1) data augment: optimized mosaic, added image random brightness processing, and limited the scaling ratio to 1–3. (2) Added a channel attention mechanism in backbone. Extract information that is more important to the task objective from numerous feature information. The efficiency and accuracy of model processing can be improved. The channel attention mechanism has been proven to use more attention resources to acquire high-value information and compress useless information (Woo et al., 2018). (3) The neck adopts the BiFPN (Bi-directional Convolutional Block Attention Module) structure. It uses learnable weights to learn the importance of different input features, repeatedly applying top-down and bottom-up multi-scale feature fusion (Tan et al., 2020).

The YOLOX series includes YOLOX-s, YOLOX-m, YOLOX-l, YOLO-x and YOLOX-Darknet53, the model size (parameters) and accuracy (usually expressed in mean average precision, mAP) increase in turn under the same conditions. Larger size model, greater arithmetic power required. Floating point operations (FLOPs) can be used to measure the complexity of the model, the larger FLOPs need more arithmetic power. The mAP (0.5) of YOLOX-m is 15.80% higher than that of YOLOX-s. YOLOX-l, YOLOX-x and YOLOX-Darknet53' mAP are only 5.97%, 8.96%, and 1.71% higher than YOLOX-m, respectively, but the FLOPs and the parameters of them are much higher than YOLOX-m (**Figure 3**). Considering the task scenario and hardware requirements, combined with the determination of YOLOX accuracy, the number of parameters, and arithmetic power of each model, YOLOX-m has high prediction accuracy, small parameters, and low computational overhead compared with other YOLOX series model. Finally, we selected the YOLOX-m model for optimization and testing.

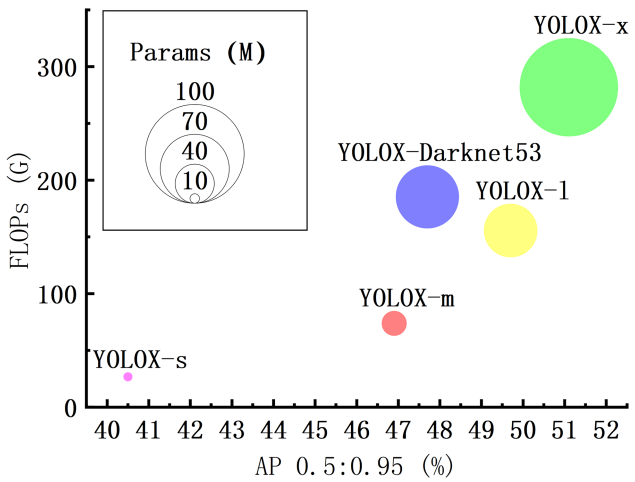


FIGURE 3 | Accuracy, number of parameters, and computational overhead of each YOLOX model.

Based on YOLOX-m, this paper proposes the following improvements for low-resolution, dense target scenes (the improved YOLOX-m network framework is shown in **Figure 4**).

Data Augmentation and Mosaic Optimization

This paper describes a modified mosaic method for data augmentation and expansion. A mosaic is four images stitched together into a new image after random changes, such as flipping and scaling while processing the labels corresponding to the target objects (Yun et al., 2019; **Figure 5**). Experiments have shown that mosaic enhancement in model training is easier to detect in small targets, such as wheat ears (Kisantal et al., 2019). Considering the low resolution of our dataset and the small, dense nature of wheat targets, a light random variation code and a restricted mosaic scaling index were added to improve the network learning. We made the following improvements to the algorithm: (1) by converting RGB to HSV, setting the boosted V-segment value, and then converting the result to RGB, we changed the image's brightness to simulate random changes in lighting; and (2) we limited the scale parameter for the mosaic in 1:3, i.e., instead of shrinking the image, we randomly zoomed-in up to 3x. For low-resolution and small target objects, the input network improved considerably over the original image after zooming-in.

In addition, the mixup algorithm was also used, which first read an image to scale up to a 640×640 image while calculating the scaled annotation frame. Next, a randomly selected image was also filled and scaled to 640×640 pix, and the scaled label box was calculated. The fusion factor was set, and the two images after the change were weighted and fused to finally obtain a mixup image (Zhang et al., 2017), where the annotation frames of the two images exist superimposed.

Backbone Added the Channel Attention Mechanism

The innovation of the Squeeze-and-Excitation Layer (SElayer) network focuses on the relationship between channels, with the

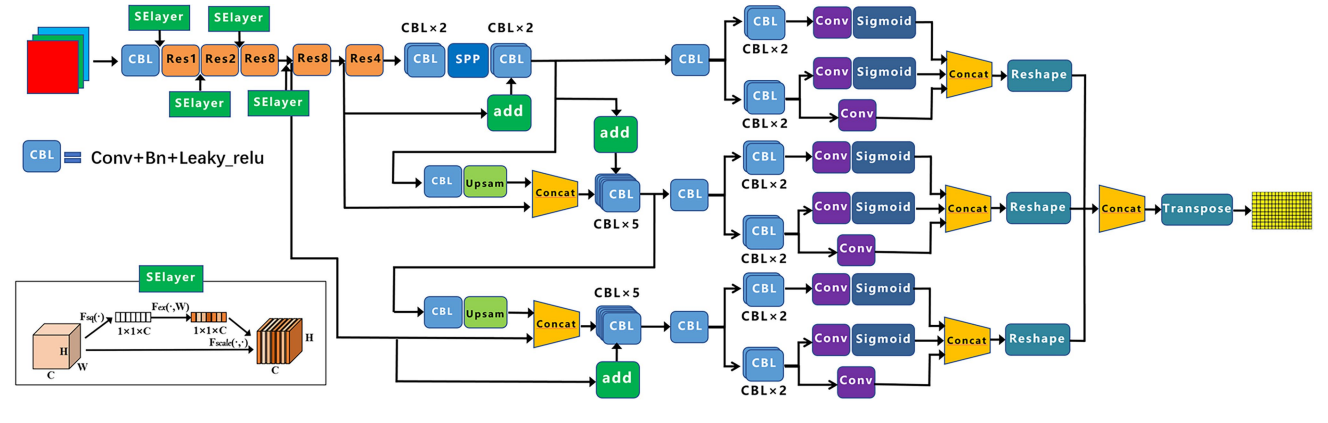


FIGURE 4 | Improved YOLOX-m framework diagram.

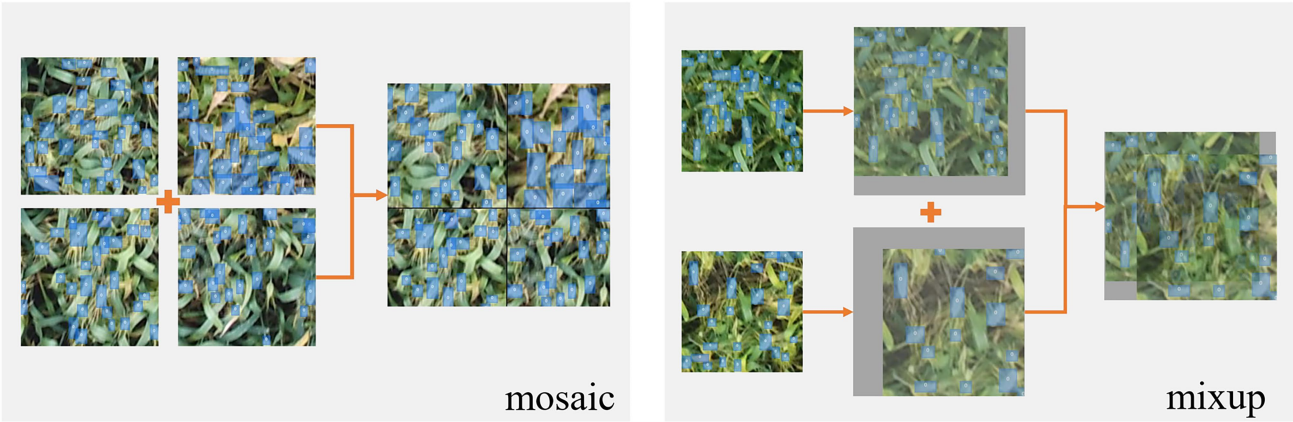


FIGURE 5 | Schematic diagram of mosaic and mixup processing.

aim for the model to automatically learn the importance of different channel features (Hu et al., 2018). In essence, convolution is the fusion of features over a local region, including the spatial (H and W dimensions) and inter-channel (C dimension) fusion of features. Small targets have weak feature representation on their own, and detection in which more feature information needs to be learned requires deepening the network structure. The channel attention mechanism allows the neural network to focus on the channels more relevant to the target task and achieve a reasonable weight distribution. Extract information that is more important to the task objective from numerous feature information, it can improve the efficiency and accuracy of model processing. The channel attention mechanism has been proven to use more attention resources to obtain high-value information and compress useless information.

Figure 4 shows the SElayer structure, which uses global average pooling. The feature map of c channels, $H \times W$, is compressed into C channels, 1×1, by Equation (7); $1 \times 1 \times C$ contains global spatial information and is compressed into a channel descriptor (channel descriptor).

$$F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{1}$$

The result of the global pooling of squeezed channels (which can be considered a C-dimensional vector) is fully connected to obtain a C/r-dimensional vector, which Relu activates. It is then fully connected again to change the C/r-dimensional vector back to a C-dimensional vector and is finally activated by a sigmoid function so that values lie between 0 and 1. This is the obtained weight matrix.

$$F_{ex}(z,W) = \sigma\left(g(z,W)\right) = \sigma\left(W_2 \delta\left(W_1 z\right)\right) \tag{2}$$

The SElayer is added to layers 1, 2, 3, and 4 after Conv + Bn + Leaky_relu (CBL). The SE module is designed mainly to improve the model's sensitivity to channel features. The module is lightweight and can be applied to existing network structures to improve performance, with only a small increase in computation.

The Neck Adopts the BiFPN Structure

YOLOX uses YOLOv3 with added SPP components as the benchmark network and Neck as the FPN structure. We added the BiFPN structure to the network, as shown in **Figure 4**. The target detection task for small objects is difficult because large objects occupy many pixel points, but small objects have few. In the convolution process, as the convolution goes deeper, the features of large objects are readily retained, while the features of small objects are easily ignored after multiple convolutions. Therefore, the FPN structure is generated, which fuses the detailed information of the lower layers and the semantic information of the higher layers, thus increasing the perceptual field of the lower layers and enabling the lower layers to obtain more contextual information when performing small object detection (Tan et al., 2020). BiFPN is a weighted bi-directional feature pyramid network that allows fast, straightforward multi-scale feature fusion to pursue a more efficient multi-scale fusion.

Evaluation of the Model Performance

The validation set in the respective dataset is used as a reference to evaluate the accuracy of the model prediction. The following metrics are selected in this paper to measure the accuracy of the model.

IOU Loss

In the IOU evaluation criteria, the L1 loss and L2 loss are obtained by summing the four coordinates of the bounding box after finding the losses separately, which ignores any correlation between the coordinates. However, the evaluation does need to consider the correlation between the coordinates. The calculation formula is as follows:

$$U\ loss = -\ln\frac{Intersection\left(Prediction, Ground\ truth\right)}{Union\left(Prediction, Ground\ truth\right)} \quad (3)$$

where Ground truth is the true frame consisting of $\tilde{x}_t, \tilde{x}_b, \tilde{x}_l$, and $\tilde{x}_r$, and Prediction is the prediction frame consisting of x_t, x_b, x_l, and x_r. The IOU loss is obtained by evaluating $-\ln(\mathrm{IOU})$ after determining the IOU. Relative to L2 loss, the IOU loss increases with the number of iterations with lower loss, and the prediction frame is more accurate (Yu et al., 2016).

Average Precision (AP50)

To assess the accuracy of the network, we tested AP50. AP50 is the average precision when the IOU of the prediction frame and that of the real frame are greater than 0.5. A higher AP means that the accuracy of the network is higher. The formula for AP is as follows:

$$AP = \sum_{k=1}^{N} \max_{\tilde{k} \geq k} P\left(\tilde{k}\right) \Delta Re\left(k\right) \quad (4)$$

where P is the accuracy rate (Equation 5), and Re is the recall rate (Equation 6).

$$P = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$\mathrm{Re} = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

True positive (TP) represents that the samples are predicted correctly and are actually positive. False-positive (FP) represents that the samples are predicted to be positive but are actually negative. In addition, False-negative (FN) represents that the samples are predicted to be negative but are actually positive.

Frame Per Second

The number of frames per second (FPS) is an important indicator to examine the real-time performance of the model. An adequate FPS can meet the demand in practical applications.

RMSE and R^2

In addition, metrics such as root mean square error (RMSE) and coefficient of determination (R^2) are used to evaluate the wheat head counting performances. The lower RMSE and higher R^2, the better performance of the model. Their counting equations are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(Truth_k - Predicted_k\right)^2} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{=1}^{n}\left(Truth_k - Predicted_k\right)^2}{\sum_{k=1}^{n}\left(Truth_k - Truth_k\right)^2} \quad (8)$$

For the improved model, we conducted ablation tests and performed statistical analysis on three series of datasets. Performance tests were conducted using the improved YOLOX-m model for different periods of wheat-ear scenes (P1–P3), different density wheat-ear scenes (D1–D3), and different resolution scenes (H1–H3).

RESULTS AND ANALYSIS

Ablation Experiments With Improved YOLOX-m Model

We randomly selected 350 annotated datasets from all annotated images for the mixed dataset test and divided the training and validation sets in the ratio of 8:2. Using the original YOLOX version of YOLOX-m as the baseline, we tested three optimization schemes: data augmentation to improve mosaic, adding SElayer, and using BiFPN. The platform configuration used for the ablation experiments used the Intel(R) Xeon(R) CPU E7-8880 v4 (2.20Ghz) × 4, RAM: 256 GB, GPU: Quadro RTX 5000 with 16 GB of video memory, CUDA version 10.0, and cudnn version 7.4. Other experiments in the following are also based on this platform. The training epoch for all models was 300 iterations, and the batch size was 6.

Compared with the standard YOLOX-m, the improved YOLOX-m-based method had the highest accuracy with an AP50 of 86.34% (**Table 2**), which was 2.74% higher than that of the standard YOLOX-m, and a speed of 40.16 FPS, which could achieve the task of wheat spike detection accurately. The standard YOLOX-m reached a high point and converged faster in the early stage. The model emerged with a larger fluctuation early after the data enhancement optimization was turned on. The fluctuation enhanced sequentially after the Attention and BiFPN were turned on, and both gradually converged after the 150th epoch (**Figure 6**). Finally, the models with standard YOLOX-m and data-enhanced optimization enabled maintained a flat trend until the end of the training. In contrast, the model with Attention enabled showed an upward change and then a downward change after the 250th epoch, and the model with BiFPN enabled showed a continuous upward trend after the 250th epoch.

Performance of the Improved Model on Different Scenario Datasets

The training results of the improved model on three series of datasets are shown in **Figure 7**. Comparing the test results on the original YOLOX-m network for a total of nine datasets

TABLE 2 | Accuracy and performance of ablation experiments with the improved YOLOX-m model.

Model improvement	AP50 (%)	FPS
YOLOX-m	84.04	39.86
+AUG	84.69 (**+0.65**)	40.37
+Attention	85.89 (**+1.20**)	40.53
+BiFPN	86.34 (**+0.45**)	40.16

The bold values means the difference between this value and the previous value.

in three series, the accuracy (AP50) of the improved model were all improved to different degrees (**Table 3**). The highest increase on the D2 dataset, which increased by 3.1%. P2 dataset increase 2.54% and the lowest increase on the H3 dataset, which increased by 0.78%. The improved YOLOX-m model can increase wheat recognition accuracy of UAV orthophotos under different practical scenarios in large fields. The following is a detailed analysis of three scenarios:

Scenario P: As shown in **Figures 8A,D**, the model performed best on the P2 dataset, with the AP50 quickly reaching the highest level and converging at the 180th epoch with a maximum AP of 88.03%. This is related to the strong contrast displayed by the wheat ears and leaves in the field during the P2 lactation period when the ears were grayish-white, differing significantly from the light green color presented by the leaves. Although slightly weaker than in the P1 flowering period, the P2 wheat-ear texture characteristics were significantly stronger than in the P3 maturity period (**Figure 7A**). The model showed a lower AP than P2 in the P1 flowering period, with a maximum AP of 80.70%, and the training curve was low at the beginning and then gradually increased, leveling off at 210 epochs, 7.33% lower than the best AP of P2. The color of wheat ears in the P1 flowering period was similar to that of the leaves, and the stacking of labeled boxes was slightly higher, indicating that the stacking of wheat ears in this period was more serious than in P2, which had an impact on the recognition of the model. The worst performance of the model was in the maturity period of P3, with the highest AP of 77.79%, 10.24% lower than the best AP of P2. The training curve started moderately, and the subsequent growth was slow, converging at 160 epochs and improving slightly, slowing down at the 270th epoch. The dataset statistics show that the wheat ears and leaves were green in this period. They are more similar, and the labeling frame stacking degree was up to 9.86%. The wheat ears are stacked to a high degree. These are the main reasons for the poor training accuracy of the model.

Scenario D: The higher the planting density, the denser the wheat ears in the same field of view of the camera, and the more severe the overlap. The model achieved 87.59% AP on the D1 dataset (**Figures 8B,E**), and the training curve converged early and fast. The D1 dataset had the lowest labeled frame overlap among all the datasets, at 2.98%, which is a more desirable dataset. Model training AP on the D2 dataset was 84.94%. With the increase of wheat planting density, the overlap between wheat ears and leaf shading gradually increased (**Figure 7B**), 2.65% compared with D1. The highest AP was

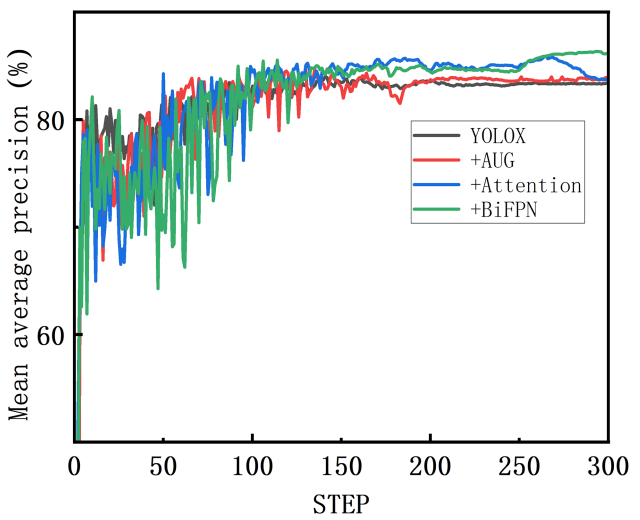


FIGURE 6 | Ablation test of YOLOX-m improved model on mixed dataset.

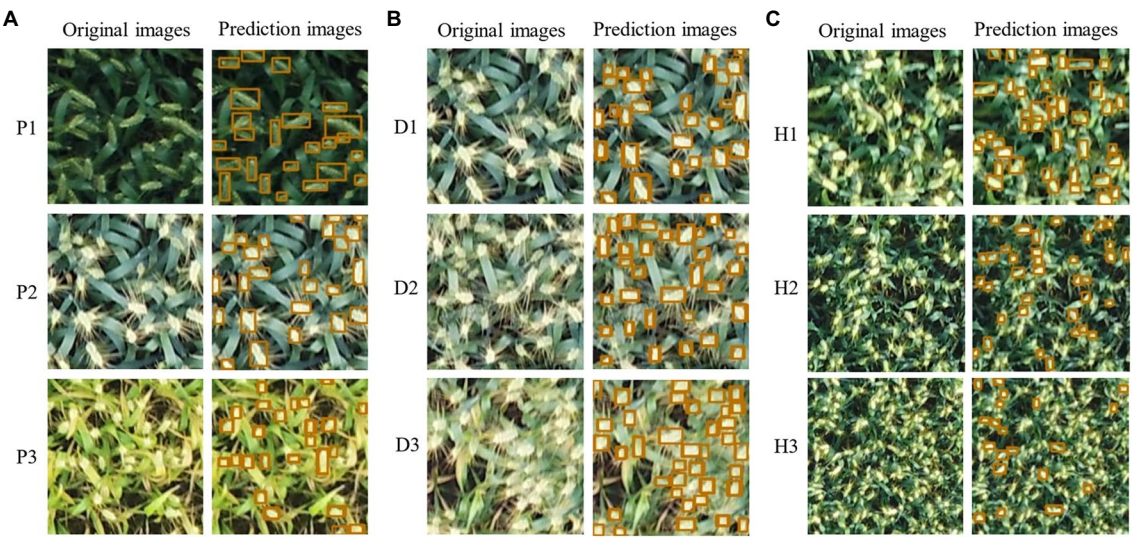


FIGURE 7 | Prediction effect of the improved model on the three series datasets: **(A)** P series, **(B)** D series, and **(C)** H series.

TABLE 3 | Comparison of accuracy and IOU loss between original and improved networks.

Dataset	Origin model		Improved model	
	AP50 (%)	IOU loss	AP50 (%)	IOU loss
P1	78.58	2.25	80.70 (+2.12)	1.18
P2	85.49	2.25	88.03 (+2.54)	1.24
P3	75.93	2.31	77.79 (+1.86)	1.33
D1	85.70	2.29	87.59 (+1.89)	1.42
D2	81.83	2.33	84.94 (+3.11)	1.59
D3	73.51	2.68	76.23 (+2.72)	1.26
H1	85.78	2.23	87.93 (+2.15)	1.28
H2	72.16	2.36	74.35 (+2.19)	1.27
H3	62.65	2.48	63.43 (+0.78)	1.33

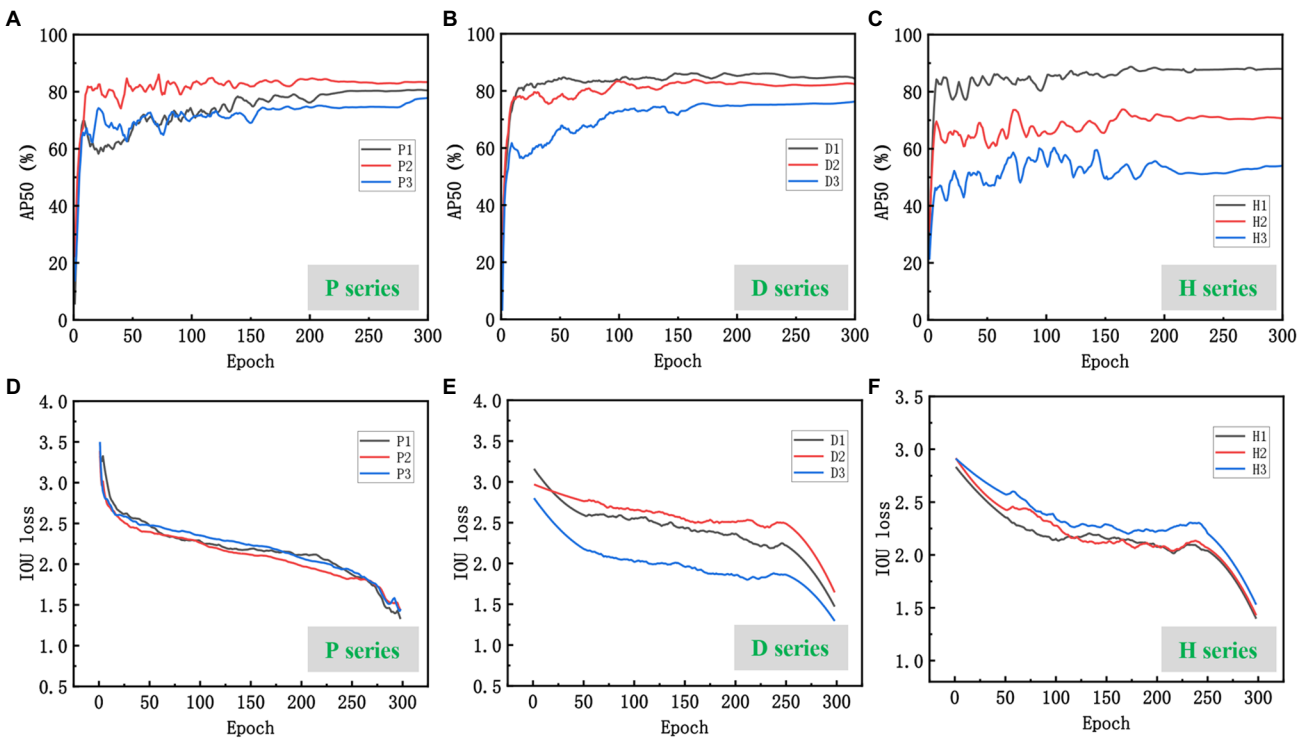


FIGURE 8 | Performance of the improved model on the three series datasets: **(A–C)** are AP50 curves for the three series datasets; **(D–F)** are IOU loss curves for the three series datasets.

only 76.23% in the D3 dataset with the highest test density, 11.36% lower than D1. Thus, the difficulty of wheat-ear recognition by UAV images increased sharply in wheat fields planted at high density.

Scenario H: The different flying heights of the UAVs affected the resolution of the images and the size of the target. And the size of the images intercepted in the three datasets in the H series experiment was 200×200 pixels (**Figure 7C**). The difference in the target object size in the H series dataset affects the model's accuracy for wheat ear's feature extraction and recognition. The larger target object, the more pronounced the target features extracted by the model and the higher the training and recognition accuracy. As shown in **Figures 8C,F**, the accuracy of the training curves of H1, H2, and H3 decreases in order, and the model performs best on the H1 dataset with the highest AP50 of 87.93% and the smallest fluctuation of the pre-training curve among the three periods and starts to converge first (about the 150th epoch); H2 has the second-highest training accuracy with the highest AP of 73.35% and starts to converge at the 180th epoch attachment. H3 has the worst effect, with the highest AP50 of only 63.43% and the most drastic fluctuations, there is a decline at 210 epochs. Then a slight upturn, and more wheat sheaves failed to be recognized by the model, as can be seen in the prediction effect graph.

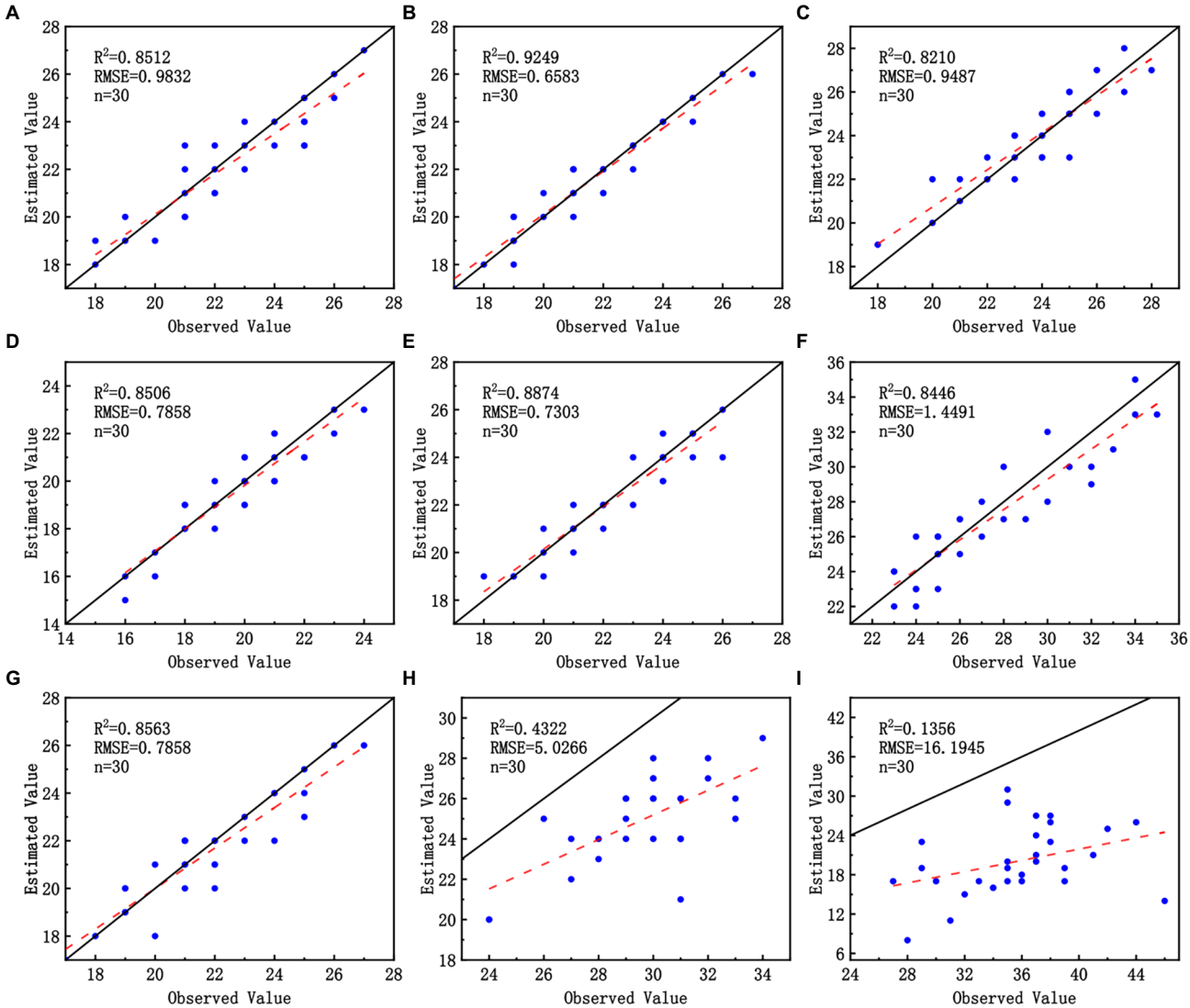


FIGURE 9 | Performance of the improved model on three wheat scenes datasets: **(A–C)** correspond to P1, P2, and P3; **(D–F)** correspond to D1, D2, and D3; **(G–I)** correspond to H1, H2, and H3.

Overall, combining the results of the three scenarios, the best UAV orthophoto recognition can be obtained from the wheat milk stage (P2), low planting density (D2), and low flight altitude (H1).

Model Counting Accuracy Validation

In each validation set, 30 images were randomly selected for manual recognition of wheat ears and counting. The results of manual observation of wheat-ear number and the recognition of wheat-ear number by the improved model were compared. The R^2 and RMSE were calculated by linear fitting (**Figure 9**). Our model showed excellent wheat-ear recognition ability on all except H2 and H3 datasets, with R^2 greater than 0.8446 and RMSE less than 1.4491. The best performance on the P2 dataset with an R^2 of 0.9249 and an RMSE of 0.6583. And the model performed poorly on the H2 and H3 datasets due to the effect of the UAV flight height on the image GSD, which resulted in the wheat ears occupying too few pixels in the image; the features were difficult to be captured by the network, which also verifies that low GSD and small targets are difficult to identify with deep networks (Zhang et al., 2020).

DISCUSSION

Analysis of Dataset Metrics

We counted and compared several metrics of the P, D, and H series datasets, including the average number of labeled boxes per figure (V_{lab}), the stacking degree of labeled boxes (D_{lab}), the average pixel of labeled boxes (V_{pix}). D_{lab} can

approximate the degree of wheat stacking in the dataset. The equations of V_{lab}, D_{lab}, and V_{pix} are as follows:

$$V_{lab} = \frac{N_{lab}}{N_{img}} \tag{9}$$

$$D_{lab} = \frac{lab_i \cap lab_j}{lab_i \cup lab_j} \tag{10}$$

$$V_{pix} = \frac{\sum S_{pix}}{N_{lab}} \tag{11}$$

where lab_i and lab_j are any two annotation frames, and S_{pix} is the sum of pixels occupied by all annotation frames.

Overall, as shown in **Table 4**, the number of images in each segmented scene dataset ranged from 255 to 395. The number of annotated boxes in each dataset was counted, with the least number of annotated boxes in the D1 dataset and the most in the P2 dataset. The average number of annotations reached a maximum of 36.0 and a minimum of 20.5.

In the P series scenes, the average number of labeled frames per image in the P1 dataset was 23.1, with 5.22% of the frames having a stacking degree greater than 0.25 and an overall dark green color. Labeled frames per image in the P2 dataset were 24.6, with 3.97% of the frames having a stacking degree greater than 0.25 and an overall light green color. The average number of annotated frames per image in the P3 dataset was 23.9, with 6.86% of the annotated frames having a stacking degree greater than 0.25; they too were yellow overall. To make the network more successful in detecting wheat ears, it was necessary to expose the ears fully. Hence, they were visible in the images, with minimal leaf shading and overlap between ears. For different genotypes and environmental conditions, we observed wheat ears with different morphologies, sizes, and strain distributions. For example, in the case of Yangmai23, we observed that the wheat tended to bend during the seed filling stage, which increased the overlap between heads. However, in the stage between tasseling and flowering, some wheat spikes were not yet fully grown and were difficult to see. The wheat fertility stage affected the wheat plant morphology and thus the differences in the angle, overlap, and color of the wheat ears in the images taken by the UAV, and P2 was a period where better identification could be obtained. Therefore, we recommend acquiring images after flowering when the wheat ears are fully emerged and still upright.

Among the D series scenes, the average number of labeled frames per image in the low-density D1 scene was 20.5, and 2.68% of the labeled frames have a stacking degree greater than 0.25, with few labeled frames stacked. The average number of labeled frames per image in the D2 dataset was 24.2, and 5.95% of the labeled frames have a stacking degree greater than 0.25. The D3 dataset was the densest, and the overlap of labeled frames greater than 0.25 reached 7.47%. The density of wheat crop planting affected the recognition accuracy of the deep network; excessive density, serious stacking occurred, and the recognition effect decreased.

In the H series scenes, the average number of annotation frames per image increased with height, and the average size of wheat-ear annotation frames decreased with height, with 1123.6 pixels for H1, 638.2 pixels for H2, and 357.6 pixels for H3. Thus, the too-small size of wheat-ear annotation frames, i.e., the size of the pixels occupied by wheat ears, affected the recognition accuracy of the depth network.

Constraint of Drone Flight Height

UAV orthophoto stitching needs to meet the synergy between parameters such as flight altitude, heading overlap rate, and side overlap rate. The UAV flight altitude is often set very low to obtain higher ground resolution (GSD) images. At too low a flight altitude, the strong wind from the UAV propeller blows the wheat plants about, making the wheat-ear tilt and swing and resulting in blurred photos and failed orthophoto stitching. This can prolong the mission time, and the data to be stored grow exponentially, requiring high UAV range and storage space. From the formula for GSD (Equation 12), it can be seen that the only condition that determines the GSD on a fixed focal length UAV is the UAV flight altitude, and too high a UAV flight altitude makes the GSD of the target object wheat ears too small for accurate identification. We tested the flight time, the number of photos, and data size required for different UAV flight heights and measured the GSD at different heights, as shown in **Table 5**. Taking Insprie2 with X5S and 45 mm fixed focal length lens as an example, setting the same heading overlap rate and side overlap rate, the lowest height that can achieve the route shooting task is 16 m, the flight time required for this task is 1.5 times as much as that of 20 m flight height task. A higher flight

TABLE 4 | Statistical table of dataset indicators.

Dataset	P Series			D Series			H Series		
	P1	P2	P3	D1	D2	D3	H1	H2	H3
Average number of annotations (pcs)	23.1	24.6	23.9	20.5	24.2	30.4	24.5	30.0	36.0
>0.25 Stacking degree share (%)	5.22	3.97	6.68	2.68	5.95	7.47	3.42	3.15	3.67
Average annotation frame pixels (pixels)	984.5	1096.4	1164.3	930.4	1013.5	1087.1	1123.6	638.2	357.6

TABLE 5 | Parameters corresponding to different flight altitudes of UAVs.

Flight altitude (m)	Heading overlap rate (%)	Sideways overlap rate (%)	GSD (cm)	Flight time (s)	Number of photos (pcs)	Data size (Mb)
15	78	80	-	-	-	-
20	78	80	0.3	569	241	2048.5
25	78	80	0.4	367	155	1317.3
30	78	80	0.5	263	110	935.7

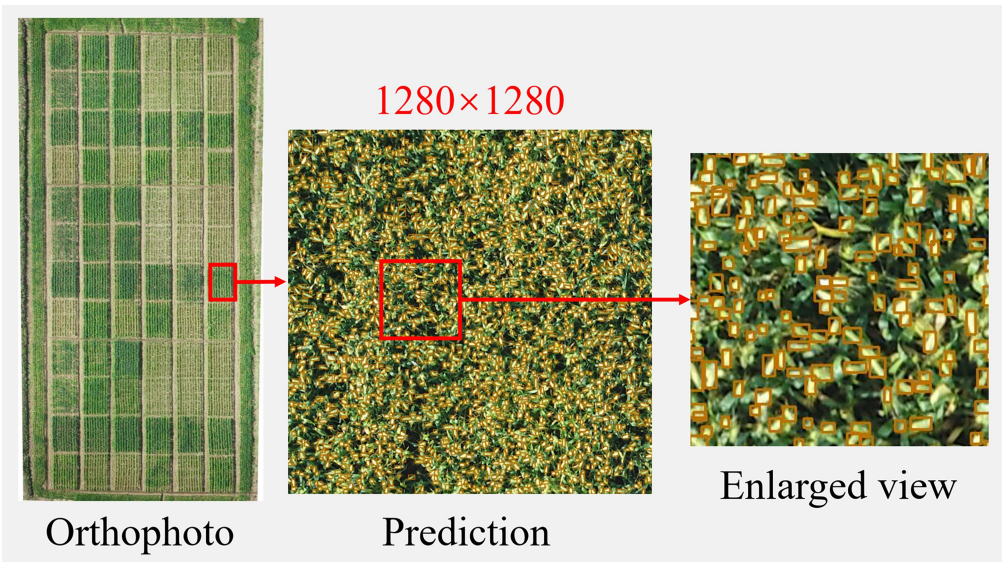


FIGURE 10 | Orthophoto cropping recognition effect.

altitude can significantly reduce the number of photos and data size, and shorten the task execution time, while reducing the GSD.

$$\mathrm{GSD} = \frac{H \times a}{f} \tag{12}$$

where H is the relative altitude, f is the focal length of the lens, and a is the image element size.

The experiment found that the ground station software could not generate operational tasks at 15 m altitudes. Combining the completed three height scenarios GSD, and the results in sections "Ablation Experiments With Improved YOLOX-m Model, Performance of the Improved Model on Different Scenario Datasets, and Model Counting Accuracy Validation" shows that the UAV flight height has a significant impact on the model results, and it is necessary to ensure that the wheat is not affected by the wind while maximizing the shooting of a large target of small wheat ears under the limitations of UAV hardware. Using a 45 mm telephoto lens, a clear image (H1) was captured at a UAV altitude of 20 m, and a better model effect was achieved. Optimizing the UAV image acquisition method in balancing the relationship between flight altitude, mission time, and GSD, we found the best solution for the current hardware equipment conditions to improve the quality of UAV orthoimages, improve the characteristics of small wheat targets, and eliminate the imbalance of data by exhausting the construction of large-scale datasets and testing different sample distributions.

Improving the Efficiency of UAV Recognition of Wheat Ears Using Larger Size Images

Finally, we tried to crop a random image of 1,280 × 1,280 pix on the UAV orthophoto and feed it into the improved model for prediction, as shown in **Figure 10**. The wheat ears could be recognized more accurately at this resolution, which indicates that the optimized network had a stronger perceptual ability and adapted to recognition at larger resolutions. The number of wheat ears in the manual statistics image was 1,689, and the network recognized 1,597, with an error of 5.45%. This method can reduce the segmentation of the UAV orthophoto processing into too many small images, which is very helpful in reducing the computation and image-processing times.

Future Trends

Recent years have seen the research on UAV detection of wheat sheaves develop rapidly. However, it is still in the research stage, and more cost is needed to make this technology applicable to practical production. In addition, it has been found that thermal infrared images have better contrast than RGB images. The difference in temperature between wheat ears and other parts of the plant is used to segment the ears and delineate the color threshold for counting (Fernandez-Gallego et al., 2019). With the development of 3D technology and the popularity of 3D devices, 3D technology has also been applied to wheat-ear identification (Velumani, 2017; Ghahremani et al., 2021), using 3D laser point cloud segmentation technology to achieve wheat-ear identification, which provides another new idea for wheat-ear counting.

CONCLUSION

This paper improved the YOLOX network by optimizing mosaics, adopting the BiFPN structure, and adding an attention mechanism. The ablation test showed that the change improved the network performance. Tests in three periods, at three densities, and for three height scenarios showed that our model had excellent wheat ear recognition on P1–P3, D1–D3, and H1 datasets, with R^2 greater than 0.8446 and RMSE less than 1.4491, and the best performance on the P2 dataset with an R^2 of 0.9249 and an RMSE of 0.6583. In comparison, H2 and H3 indicate that deep network recognition is difficult under the condition of low GSD. We suggest acquiring images after flowering when the wheat ears have fully emerged and are still upright. This sets up the best UAV flight plan with hardware devices to improve the quality of UAV orthoimages for the best training and recognition results.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YZ and LT: conceptualization. YZ: methodology, formal analysis, data curation, and writing—original draft preparation. JC: software. YT and JC: validation. SC: writing—review and editing and funding acquisition. LT: project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This research was financially supported by the National Natural Science Foundation of China (32172110, 32001465, and 31872852), the Key Research and Development Program (Modern Agriculture) of Jiangsu Province (BE2020319), and the Postgraduate Research Practice Innovation Program of Jiangsu Province (KYCX19_2107).

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

REFERENCES

Cointault, F., and Gouton, P. (2007). "Texture or color analysis in agronomic images for wheat ear counting." in *2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System*. December 16-18, 2007.

David, E., Serouart, M., Smith, D., Madec, S., Velumani, K., Liu, S., et al. (2021). Global wheat head detection 2021: an improved dataset for benchmarking wheat head detection methods. *Plant Phenomics* 2021:9846158. doi: 10.34133/2021/9846158

Fernandez-Gallego, J., Buchaillot, M., Aparicio Gutiérrez, N., Nieto-Taladriz, M., Araus, J., and Kefauver, S. (2019). Automatic wheat ear counting using thermal imagery. *Remote Sens.* 11, 751–764. doi: 10.3390/rs11070751

Fernandez-Gallego, J. A., Lootens, P., Borra-Serrano, I., Derycke, V., Haesaert, G., Roldan-Ruiz, I., et al. (2020). Automatic wheat ear counting using machine learning based on RGB UAV imagery. *Plant J.* 103, 1603–1613. doi: 10.1111/tpj.14799

Frdric, C., Ludovic, J., Gilles, R., Christian, G., David, O., Marie-France, D., et al. (2012). "Texture, color and frequential proxy-detection image processing for crop characterization in a context of precision agriculture," in *Agricultural Science.* ed. A. Godwin (INTECH).

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv [Preprint].

Ghahremani, M., Williams, K., Corke, F. M. K., 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

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks." in *Proceedings of the IEEE conference on computer vision and pattern recognition*, August 20-24, 2018; 7132–7141.

Khaki, S., Safaei, N., Pham, H., and Wang, L. (2021). Wheatnet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. arXiv [Preprint].

Kisantal, M., Zbigniew, W., Jakub, M., Jacek, N., and Kyunghyun, C. (2019). Augmentation for small object detection. arXiv [Preprint].

Li, J., Li, C., Fei, S., Ma, C., Chen, W., Ding, F., et al. (2021). Wheat ear recognition based on retinanet and transfer learning. *Sensors* 21, 4845–4866. doi: 10.3390/s21144845

Misra, T., Arora, A., Marwaha, S., Chinnusamy, V., Rao, A. R., Jain, R., et al. (2020). SpikeSegNet-a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* 16–40. doi: 10.1186/s13007-020-00582-9

Panboonyuen, T., Thongbai, S., Wongweeranimit, W., Santitamnont, P., Suphan, K., and Charoenphon, C. (2021). Object detection of road assets using transformer-based YOLOX with feature pyramid decoder on thai highway panorama. *Information* 13:5. doi: 10.3390/info13010005

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., and Hawkesford, M. J. (2019). DeepCount: in-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks. *Front. Plant Sci.* 10:1176. doi: 10.3389/fpls.2019.01176

Sun, Y., Chen, Y., Jin, X., Yu, J., and Chen, Y. (2021). AI differentiation of bok choy seedlings from weeds. *Fujian J. Agric. Sci.* 36, 1484–1490. doi: 10.19303/j.issn.1008-0384.2021.12.013

Tan, M., Pang, R., and Le, Q.V. (2020). "EfficientDet: scalable and efficient object detection." in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020.

Tao, L., Chengming, S., Lijian, W., Xiaochun, Z., Xinkai, Z., and Wenshan, G. (2014). In-field wheatear counting based on image processing technology.

Trans. Chin. Soc. Agric. Mach. 45, 282–290. doi: 10.6041/j.issn.1000-1298.2014.02.047

Velumani, K. (2017). Wheat Ear Detection in Plots by Segmenting Mobile Laser Scanner Data. master's thesis. University of Twente.

Wang, Y., Qin, Y., and Cui, J. (2021). Occlusion robust wheat ear counting algorithm based on deep learning. *Front. Plant Sci.* 12:645899. doi: 10.3389/fpls.2021.645899

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)*, September 8-14, 2018; 3–19.

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., and Shen, C. (2019). TasselNetv2: in-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* 15:150. doi: 10.1186/s13007-019-0537-2

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., et al. (2020). Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant Methods* 16:106. doi: 10.1186/s13007-020-00648-8

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, 1202–1219. doi: 10.3390/agronomy11061202

Yang, J., Li, F., Sun, Z., and Jiang, S. J. J. I. H. M. S. P. (2016). A small target detection method based on human visual system and confidence. *Measurement* 7, 448–459.

Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. (2016). "Unitbox: an advanced object detection network." in *Proceedings of the 24th ACM international conference on Multimedia*. October 15-19, 2016; 516–520.

Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., and Yoo, Y. (2019). "Cutmix: regularization strategy to train strong classifiers with localizable features." in *Proceedings of the IEEE/CVF international conference on computer vision*. October 27–Novermber 2, 2019; 6023–6032.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2017). Mixup: Beyond Empirical Risk Minimization. arXiv [Preprint].

Zhang, J., Wang, C., Yang, C., Xie, T., Jiang, Z., Hu, T., et al. (2020). Assessing the effect of real spatial resolution of in situ UAV multispectral images on seedling rapeseed growth monitoring. *Remote Sens.* 12:1207. doi: 10.3390/rs12071207

Zhang, M., Wang, C., Yang, J., and Zheng, K. (2021). "Research on engineering vehicle target detection in aerial photography environment based on YOLOX." in *2021 14th International Symposium on Computational Intelligence and Design (ISCID)*. December 11-12, 2021.

Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J., and Yang, G. (2018). Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. *Front. Plant Sci.* 9:1024. doi: 10.3389/fpls.2018.01024

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 April 2022
doi: 10.3389/fpls.2022.872190

Enzyme-Free Electrochemical Sensors for *in situ* Quantification of Reducing Sugars Based on Carboxylated Graphene–Carboxylated Multiwalled Carbon Nanotubes–Gold Nanoparticle–Modified Electrode

Ke Liu[1,2,3†], Xiaodong Wang[1,2†], Bin Luo[2], Cheng Wang[2], Peichen Hou[2], Hongtu Dong[2], Aixue Li[1,2] and Chunjiang Zhao[1,2*]*

[1]Heyuan Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Heyuan, China, [2]Research Center of Intelligent Equipment, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, [3]College of Landscape and Ecological Engineering, Hebei University of Engineering, Handan, China

The reducing sugars of plants, including glucose, fructose, arabinose, galactose, xylose, and mannose, are not only the energy source of plants, but also have the messenger function of hormones in signal transduction. Moreover, they also determine the quality and flavor of agricultural products. Therefore, the *in situ* quantification of reducing sugars in plants or agriculture products is very important in precision agriculture. However, the upper detection limit of the currently developed sugar sensor is not high enough for *in situ* detection. In this study, an enzyme-free electrochemical sensor for *in situ* detection of reducing sugars was developed. Three-dimensional composite materials based on carboxylated graphene–carboxylated multi-walled carbon nanotubes attaching with gold nanoparticles (COOH-GR–COOH-MWNT–AuNPs) were formed and applied for the non-enzymatic determination of glucose, fructose, arabinose, mannose, xylose, and galactose. It was demonstrated that the COOH-GR–COOH-MWNT–AuNP-modified electrode exhibited a good catalysis behavior to these reducing sugars due to the synergistic effect of the COOH-GR, COOH-MWNT, and AuNPs. The detection range of the sensor for glucose, fructose, arabinose, mannose, xylose, and galactose is 5–80, 2–20, 2–50, 5–60, 2–40, and 5–40 mM, respectively. To our knowledge, the upper detection limit of our enzyme-free sugar sensor is the highest compared to previous studies, which is more suitable for *in situ* detection of sugars in agricultural products and plants. In addition, this sensor is simple and portable, with good reproducibility and accuracy; it will have broad practical application value in precision agriculture.

Keywords: *in situ*, enzyme-free, reducing sugars, carboxylated graphene, carboxylated multi-walled carbon nanotubes, screen-printed electrode

OPEN ACCESS

Edited by:
Daobilige Su,
China Agricultural University, China

Reviewed by:
Guzel Ziyatdinova,
Kazan Federal University, Russia
Mehmet Lütfi Yola,
Hasan Kalyoncu University, Turkey
Soo-Jin Park,
Inha University, South Korea

***Correspondence:**
Aixue Li
liax@nercita.org.cn
Chunjiang Zhao
zhaocj@nercita.org.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *09 February 2022*
Accepted: *12 April 2022*
Published: *28 April 2022*

Citation:
Liu K, Wang X, Luo B, Wang C, Hou P, Dong H, Li A and Zhao C (2022) Enzyme-Free Electrochemical Sensors for in situ Quantification of Reducing Sugars Based on Carboxylated Graphene–Carboxylated Multiwalled Carbon Nanotubes–Gold Nanoparticle–Modified Electrode. Front. Plant Sci. 13:872190. doi: 10.3389/fpls.2022.872190

INTRODUCTION

Sugars play an important role in plant growth and development. They are not only the energy source of plants, but also have the messenger function of hormones in the process of signal transduction (Li and Sheen, 2016). They also determine the quality and flavor of agricultural products. Moreover, sugar-related materials, such as cellulose, can be developed into eco-friendly and economically favorable biosorbents and biocomposites for removing some toxic substances, such as acid dye (Kamran et al., 2022), CO_2 (Kamran and Park, 2021a,b), Li^+ (Kamran and Park, 2020, 2022), and bacteria (Kamran et al., 2019a). Therefore, quantitative analysis of sugar in plants and agricultural products is very important. The traditional methods for determining sugar content include chromatography (Wahjudi et al., 2010), fluorescence method (Pablos et al., 2015), spectrophotometry (Biscay et al., 2012), and colorimetry (Ayyub et al., 2013). However, most of these methods need to be equipped with large-scale instruments and have poor portability (Li et al., 2017). With the development of precision agriculture, researchers often need to conduct *in situ* and on-site detection of the sugar content in plants or agricultural products. Therefore, there is an urgent need to develop new detection methods to achieve in situ and on-site detection of sugars in plants or agricultural products.

Electrochemical biosensor has the advantages of high sensitivity, good selectivity, good portability, fast response, and easy integration. Its development provides an effective solution for in situ and on-site measurement. In recent years, researchers have developed a variety of sugar sensors, such as glucose sensors (Gao et al., 2016; Strakosas et al., 2019) and fructose sensors (Gota et al., 2017; Xu et al., 2018). Compared with enzyme biosensors, enzyme-free biosensors have the advantages of independent of enzyme, less affected by environmental factors, low cost, good stability, and simple preparation (Xu et al., 2014a). For example, Shekarchizadeh et al. (2013) developed an enzyme-free sensor modified with copper oxide nanoparticles and multi-walled carbon nanotubes to improve its electrical activity and selectivity for the detection of glucose and fructose. de Sá et al. (2016) modified glassy carbon electrode with carbon nanotubes and metal hydroxyl oxides to detect and quantitatively analyze carbohydrates (glucose, xylose, galactose, and mannose) in sugarcane. In plants or agricultural products, the content of sugars is very high, ranging from a few millimoles to thousands of millimoles (Zhou et al., 2019). However, most developed glucose sensors are mainly used to detect glucose in humans and animals; their detection range is not suitable for plants. In addition, the highest upper detection limit of the developed sugar sensor is only more than 10 millimoles (Jeong et al., 2018), which is not high enough for *in situ* detection. Therefore, it is necessary to develop new electrochemical sensors for *in situ* detecting sugars in plants or agricultural products.

Carbon-based nanomaterials, such as graphene (GR; Hernaez, 2020), multi-walled carbon nanotubes (MWNT), carbon spherical shells (Campos et al., 2018), and carbon black (Raymundo-Pereira et al., 2017), have received extensive attention in sensor construction due to their extraordinary physical and chemical properties (Kamran et al., 2019b). GR can be obtained by chemical reduction after graphite oxidation, but it is prone to aggregate due to π–π interaction (Shim, 2019). As a special allotrope of GR, MWNT has a unique structure and performance, such as good conductivity, perfect chemical stability, acceleration ability of electron transfer on electrode surface, and large surface area (Özcan et al., 2020a,b). MWCNT- and MWCNT-based nanocomposites have a wide range of applications in the electrochemical field, such as nanosensors (Yola et al., 2021; Yola and Atar, 2021) and fuel cell (Gizem Güneştekin et al., 2020). However, if MWNT cannot be sufficiently dispersed to form a network to meet the electrical conductivity, it will not be able to get a better performance (Zhang et al., 2017). Recent studies have shown that the above problems can be effectively avoided by introducing MWNT between GR nanosheets (Su et al., 2017; Wang et al., 2018). The good dispersion of MWNT can avoid the aggregation of GR flakes. As a surfactant, GR nanoflakes can also directly disperse MWNT to form a three-dimensional (3D) network structure with large specific surface area and excellent electrical conductivity (Cui et al., 2015; Tourani et al., 2015). Moreover, the carboxylated graphene (COOH-GR) and carboxylated multi-walled carbon nanotubes (COOH-MWNT) have better hydrophilicity, biocompatibility, and carboxyl functional groups, which will further improve the sensor performance. Gold (Au) catalysts are attractive nanomaterials due to their excellent photoelectric properties and catalytic activity. Therefore, they are widely used in sensor field, such as glucose oxidation and vitamin detection (Sharma et al., 2020). This catalysis also occurs in other reducing sugars, such as fructose, arabinose, galactose, xylose, and mannose. All these reducing sugars have very similar structures, so the catalytic reaction of Au to them is similar.

Screen-printed electrode (SPE) is widely used because of its low cost, small size, mature manufacturing technology, and good electrochemical performance (Pohanka, 2020). In this study, SPE was used as the basic electrode, and COOH-GR and COOH-MWNT were used to construct a 3D network structure to immobilize AuNPs. Then, this COOH-GR–COOH-MWNT–AuNPs composite material-modified SPE was used to catalyze six reducing sugars (glucose, fructose, arabinose, galactose, xylose, and mannose), and an enzyme-free electrochemical sensor for these six reducing sugars was developed. The upper detection limit of our sensor was improved to 80 millimoles (for glucose), which will have broad application prospect in the *in situ* detection of reducing sugars in plants and agricultural products.

MATERIALS AND METHODS

Chemicals

Carboxyl graphene (GR-COOH) and carboxyl multi-walled carbon nanotube (MWNT-COOH) were purchased from Xianfeng Nanomaterials Technology Co., Ltd. (Nanjing, China). D-glucose, D-fructose, D-galactose, trisodium citrate ($C_6H_5Na_3O_7$), citric acid, and anhydrous malic acid were

purchased from Sinopharm Chemical Reagent Co., Ltd. Company (Shanghai China). Chloroauric acid ($HAuCl_4$), nafion solution (5wt%), D-xylose (xylose), L-arabinose (arabinose), mannose (mannose), D-leucine, DL-tryptophan, lysine, magnesium chloride, sucrose, Betaine, 3-indoleacetic acid, abscisic acid, gibberellins, and ascorbic acid were purchased from Sigma Reagent Co., Ltd. (St. Louis, Missouri, United States). The rest of the reagents are of analytical grade, and ultrapure water was used to prepare the solution throughout the experiment.

Instruments

Equipped with X-ray energy spectrum analysis (EDS), field emission scanning electron microscopy (FESEM) system (ZEISS SEM 500, Germany), Fourier infrared spectrometer (Thermo Nicolet IS5) were used to study the different modification steps SPE morphology. Glassy carbon sheets were used for SEM, EDS-mapping, and FTIR characterization. The detection of sugars was performed on CHI760E electrochemical workstation (Shanghai Chenhua Instrument Co., Ltd., China). Three-electrode SPEs was used, in which the working electrode and counter electrode are carbon-based and the reference electrode is silver/silver chloride. The diameter of the working electrode is 2.5 mm.

Preparation of COOH-GR–COOH-MWNT–AuNPs Composite

First, 0.5 mg/ml of COOH-GR and 1.5 mg/ml of COOH-MWNT were mixed. Fifteen milliliters of 23.6 mM $HAuCl_4$ solution was added to 20 ml of COOH-GR–COOH-MWNT mixture and stirred magnetically for 60 min. Subsequently, 20 ml of 68 mM trisodium citrate solution was added to the mixture, and the mixture was stirred magnetically for 30 min. Then, this mixture was heated at 80°C for 30 min. The resulting solution was centrifuged at 17,000 rpm for 10 min. Then the precipitate was collected and dried at 60°C for 12 h. 2.5 ml ethanol and 55 µl nafion solution were added to 55 mg dried material. Then, the COOH-GR–COOH-MWNT–AuNPs' composite was obtained. Four microliters of the COOH-GR–COOH-MWNT–AuNPs solution was used to modify the SPE electrode by dropping method. The modification process is shown in **Figure 1**.

Measurement Procedure

Cyclic voltammetry (CV) was used to study the catalytic effect of the modified electrode toward different sugars, the scanning range was −0.6 to 0.6 V, and the scanning speed was 0.05 V/s. The concentration of sugars was detected by chronoamperometry (i-t); the working voltage is 0.3 V. All electrochemical tests were carried out in 0.1 M NaOH solution.

Sugars Determination in Apple Juice by Traditional Analytical Methods

Reference control measurement of glucose and fructose concentration in apple juice was carried out by high-pressure liquid chromatography (HPLC). For arabinose, mannose, xylose, and galactose, they cannot be separated by the HPLC method. So the ion chromatography (IC) method was used to measure these sugars. The apple juice was bought from local supermarket. The apple juice was filtered through a nylon filter (aperture 0.45 µm). Glucose and fructose were determined by Agilent chromatograph. Arabinose, mannose, xylose, and galactose were determined by the ICS-3000 chromatograph. They were separated on an amino column and eluted with 78% acetonitrile solution. Glucose and fructose were detected by PA1 differential refractive index detector, and the other four sugars were detected by amperometric pulse detector.

RESULTS AND DISCUSSION

Morphology and Structure Characterization of the Sensor

Figure 2 shows the result of SEM. It can be seen from **Figure 2A** that the surface of the bare SPE is smooth without any impurities. **Figure 2B** shows the modified SPE. The lamellar wrinkled COOH-GR structure can be observed, tubular COOH-MWNT intersperse between COOH-GR nanosheets, and gold nanoparticles are distributed in the COOH-GR–COOH-MWNT. The size of gold nanoparticles is about 30–50 nm. **Figures 2C–F** shows the EDS mapping spectrum after the SPE was modified with COOH-GR–COOH-MWNT–AuNPs, and the signals of C, F, Au, and O elements are obtained. The existence of C element is attributed to the C element in COOH-GR, COOH-MWNT, and glassy carbon sheet. Since nafion contains a large amount of F element, it leads to the emergence of F element. The O elements are derived from graphene, and the Au in COOH-GR–COOH-MWNT–AuNPs is the reason for the appearance of Au element. The results of SEM and EDS proved that COOH-GR–COOH-MWNT–AuNP materials have been successfully modified on the electrode surface.

Supplementary Figure S1 shows the bare and COOH-GR–COOH-MWNT–AuNP images by FTIR. Both C=O at 1,720 cm^{-1} and O-H at 3,346 cm^{-1} are characteristic peaks of carboxyl groups, which attribute to the COOH-GR and COOH-MWNT. Glassy carbon sheet was used as the substrate for this experiment. It is well known that glassy carbon sheet comprises thin, tangled ribbons of cross-linked graphite-like sheets that share sp2 bonding and the basic structure of a six-member ring (Shi and Shiu, 2002). But the cross-linked six-member rings are possibly broken up at the surface in the polishing process. Moieties containing alcohol, phenol, aldehyde, ketone (or quinine), and carboxylic acid (or anhydride) are appeared to connect to the skeleton of glassy carbon sheet (Wu et al., 2008). Therefore, there was no significant difference between the bare and COOH-GR–COOH-MWNT–AuNP images. The effect of Au nanoparticles in FTIR spectrum cannot be observed, since they do not have molecular bonds. Similar results were also observed by Tabatabaie and Dorranian (2016) and Najafianpour and Dorranian (2018).

Electrochemical Characterization of the Sensor Preparation Process

First, the preparation process of the sensor was characterized by the CV method (**Supplementary Figure S2A**). The CV

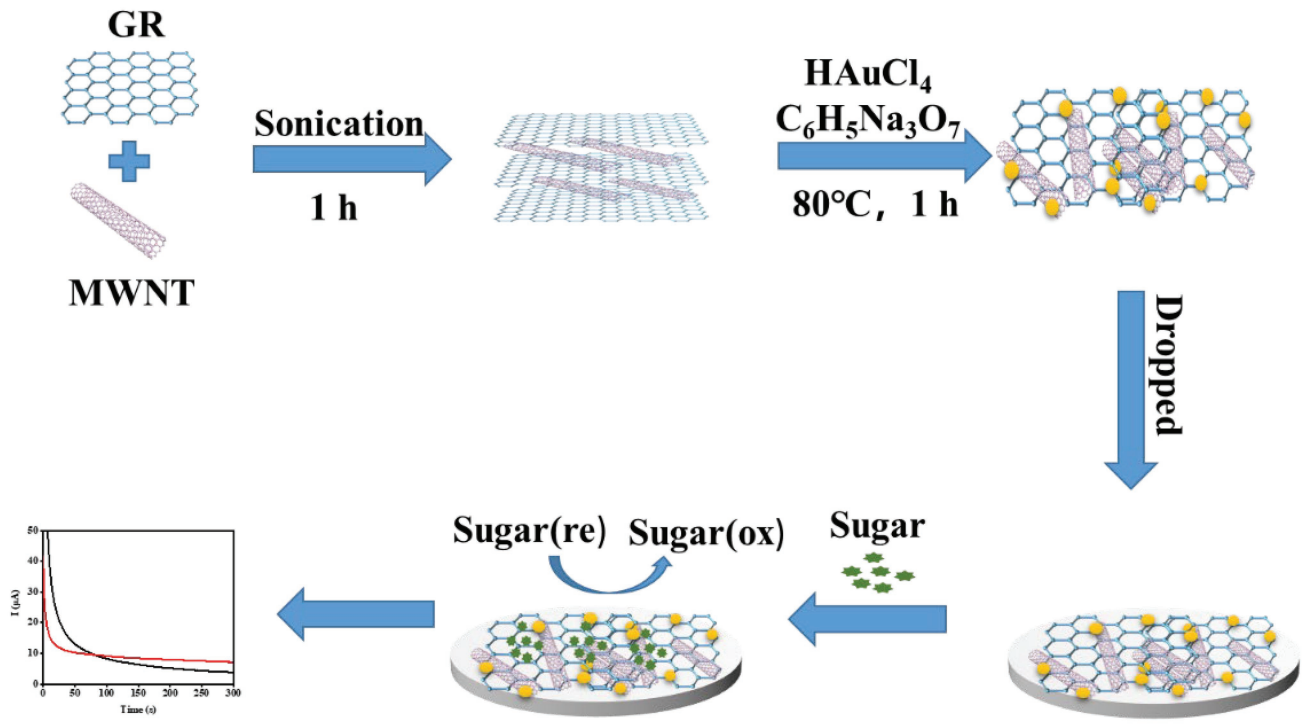


FIGURE 1 | Schematic diagram of preparation process of the enzyme-free sugar sensor.

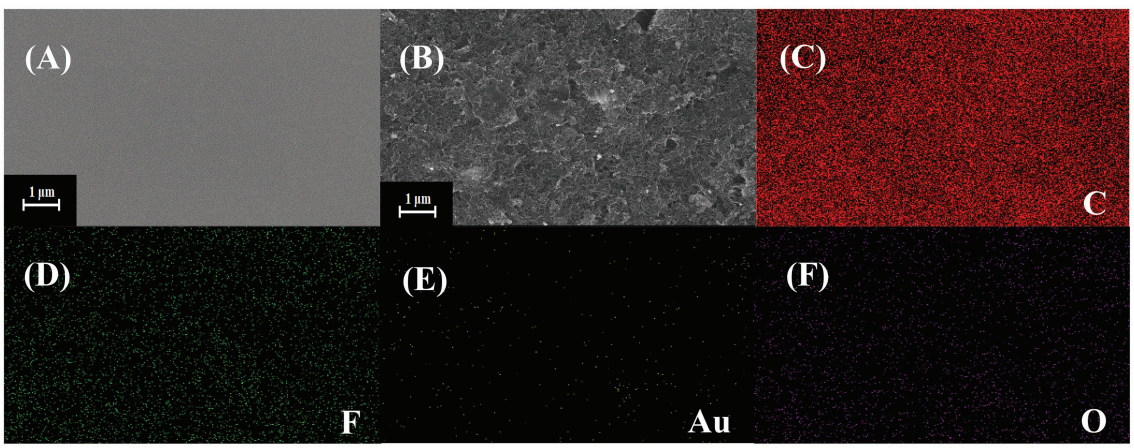


FIGURE 2 | SEM images of **(A)** bare SPE, **(B)** COOH-GR–COOH-MWNT–AuNPs/SPE, **(C–F)** are the EDS mapping images of COOH-GR–COOH-MWNT–AuNPs/SPE.

scan was performed in a 5 mM $[Fe(CN)_6]^{3-/4-}$ solution (containing 0.1 M KCL). When the COOH-GR–COOH-MWNT–AuNPs material was modified on the electrode, the redox peak current increased and the peak-to-peak potential difference decreased. This is due to that the high conductivity and catalytic performance of COOH-GR, COOH-MWNT, and AuNPs increase the electron transfer rate and enhance the reversibility of the electrode. Consistent with that of CV, the Nyquist curve (**Supplementary Figure S2B**) of the bare electrode has a smaller half arc, while there is almost no half arc after the electrode was modified with COOH-GR–COOH-MWNT–AuNPs nanocomposite. After fitting with a simple equivalent circuit model (inset in **Figure 3B**), the interfacial electron transfer resistance Rct can be obtained. The Rct value of GR-MWNT-Au/SPE (1,038 Ω) is lower than that of bare electrode (508.4 Ω), which also attributes to the high conductivity and catalytic performance of GR and MWNT. The results of CV and EIS both prove that the preparation of the sensor

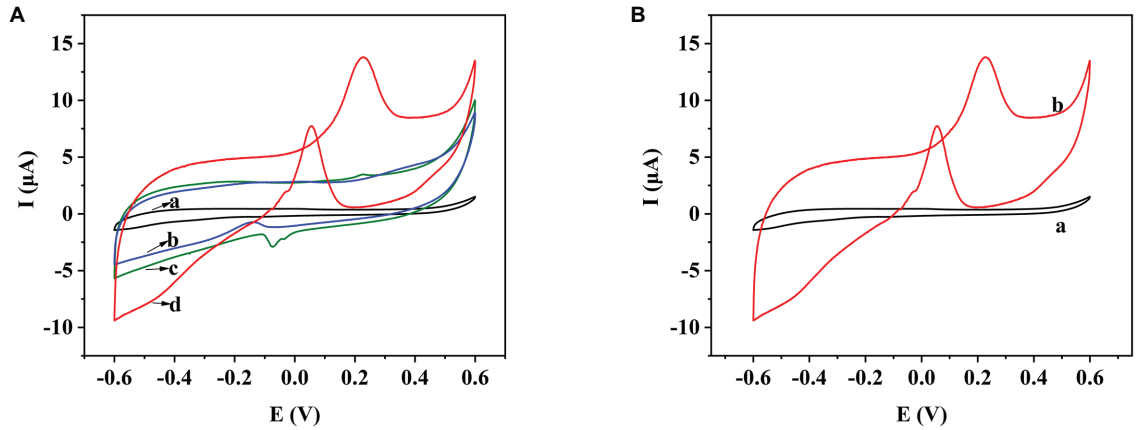


FIGURE 3 | (A) CV behavior of bare SPE (a), COOH-GR/SPE (b), COOH-MWNT/SPE (c), and COOH-GR–COOH-MWNT–AuNPs/SPE (d) in 20 mM glucose. **(B)** CV behavior of COOH-GR–COOH-MWNT–AuNPs/SPE with and without 20 mM glucose.

is successful and effective. **Supplementary Figure S2C** shows the CV graph of COOH-GR–COOH-MWNT–AuNPs/SPE in 5 mM $[Fe(CN)_6]^{3-/4-}$ solution (containing 0.1 M KCL) at various scan rates. The effective surface area of the different modified SPE was evaluated based on the Randles–Sevcik equation (Xu et al., 2014b):

$$I_P = 2.69\times10^5 \times n^{3/2} A D_0^{1/2} C v^{1/2}$$

where D_0 is the diffusion coefficient of the molecule in solution ($cm^2\ s^{-1}$), A is the effective area of the electrode (cm^2), v is the scan rate ($V\,s^{-1}$), n is the number of electrons including in the redox reaction, and C_0 is the concentration of the probe in the solution ($mol\,cm^{-3}$). For $[Fe(CN)_6]^{3-}/[Fe(CN)_6]^{4-}$, $n=1$, $C_0 = 5\times10^{-6}\,mol\,cm^{-3}$, $D_0 = 1\times10^{-5}\ cm^2\ s^{-1}$ (Wang et al., 2012). The effective surface area was 0.1041 cm^2 for the COOH-GR–COOH-MWNT–AuNPs/SPE, respectively, which was much higher than that of bare SPE (0.030 cm^2).

Electrochemical Performance of COOH-GR–COOH-MWNT–AuNP Nanomaterials

In order to test the electrochemical performance of COOH-GR–COOH-MWNT–AuNP nanomaterials, the electrochemical behaviors of bare SPE, COOH-GR–AuNPs/SPE, COOH-MWNT–AuNPs/SPE and COOH-GR–COOH-MWNT–AuNPs/SPE in the range of −0.6 to 0.6 V were investigated. Taking glucose as an example, the CV scan in **Figure 3A** was performed in 20 mM glucose (containing 0.1 M NaOH). The bare SPE electrode (curve a) does not show any oxidation peak in the range of −0.6 to 0.6 V. Because COOH-GR, COOH-MWNT, and AuNPs are all highly conductive, which can improved the electrochemical catalytic behavior of the sensor, an oxidation peak was appeared at about 0.3 V in GR-Au/SPE (curve b) and MWNT-Au/SPE (curve c), which was due to the oxidation of glucose. The highest oxidation peak was obtained on COOH-GR–COOH-MWNT–AuNPs/SPE at about 0.3 V (curve d), which indicates the synergistic effect of the COOH-GR, COOH-MWNT, and AuNPs in the catalysis of glucose. Therefore, COOH-GR–COOH-MWNT–AuNPs material was chosen for subsequent experiments. **Figure 3B** shows that there is no oxidation peak in COOH-GR–COOH-MWNT–AuNPs/SPE without the addition of glucose (curve a), while an obvious oxidation peak was observed at about 0.3 V after adding 20 mM glucose (b). This result also confirms that the appearance of oxidation peak is due to the oxidation of glucose, not the COOH-GR, COOH-MWNT, or Au nanoparticles. This catalysis also occurs in other sugars. In the presence of fructose, galactose, arabinose, mannose, or xylose, respectively, similar signals to glucose can be observed. Since these sugars have very similar structures, they are monosaccharides containing aldehyde or ketone groups, the COOH-GR–COOH-MWNT–AuNP nanomaterials have similar catalytic effect to these sugars (Xu et al., 2014a). The modified electrode can catalyze the oxidation of these sugars to form corresponding esters, which are hydrolyzed to form acids (Wang et al., 2020).

Optimization of Sensor Preparation Conditions

In order to achieve the best performance of the sensor, the conditions for preparing the sensor were optimized. In this study, the total amount of immobilized carbon nanomaterials was 2 mg/ml. The effect of single COOH-GR, COOH-MWNT, and COOH-GR-COOH-MWNT composite materials with different ratios (3:1, 2:2, 1:3) on the response current was investigated. The i-t response was measured using a glucose concentration of 20 mM. As shown in **Figure 4A**, in the various ratios of COOH-GR–COOH-MWNT composites, the maximum current response is obtained when COOH-GR: COOH-MWNT is 1:3, which shows that COOH-GR and COOH-MWNT have the best synergistic effect at this ratio. Therefore, the optimal content of COOH-GR and COOH-MWNT in this study is 0.5 and 1.5 mg/ml, respectively.

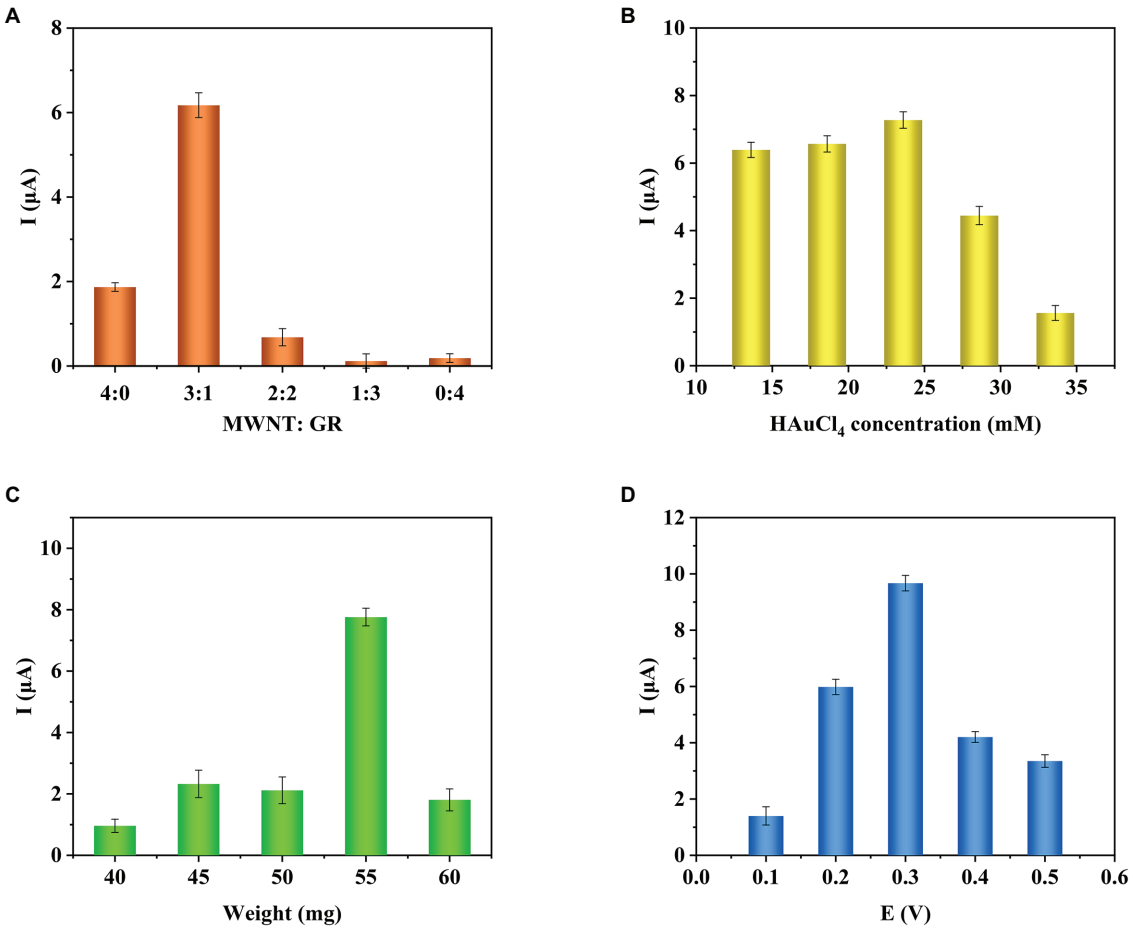


FIGURE 4 | Effect of COOH-GR–COOH-MWNT content ratio **(A)**, $HAuCl_4$ concentration **(B)**, total weight of dropped composite material **(C)**, working voltage **(D)** on the current response.

Gold nanoparticles are one of the important components of the composite material, so it is necessary to optimize the concentration of $HAuCl_4$ (13.6, 18.6, 23.6, 28.6, 33.6 mM). The optimization results are shown in **Figure 4B**. It can be seen that when the $HAuCl_4$ content is from 13.6 to 23.6 mM, the current response gradually increases. When the $HAuCl_4$ content continues to increase, the response current no longer increases, so 23.6 mM was selected for $HAuCl_4$ concentration for the subsequent experiments.

The total weight of the dropped composite material affects the performance of the modified electrode, which also requires optimization (40, 45, 50, 55, and 60 mg). The results are shown in **Figure 4C**. When the total weight of the material is 55 mg, the response current is highest. Therefore, the total weight of the composite material is determined to be 55 mg.

Finally, the influence of different voltages (0.1, 0.2, 0.3, 0.4, and 0.5 V) on the response current was examined. As shown in **Figure 4D**, when the voltage increases from 0.1 to 0.3 V, the response current gradually increases. When the voltage continued to increase, the current dropped. So 0.3 V is the best voltage for the catalytic reaction of glucose.

Performance of the Enzyme-Free Sugar Sensor

Under the optimal experimental conditions, the prepared enzyme-free sugars sensor was used to measure a series of concentrations of glucose, fructose, arabinose, mannose, xylose, or galactose solutions, respectively. The i-t curve for the detection of different concentrations of glucose is shown in **Figure 5A**. The results of linear fitting are shown in the inset of **Figure 5A**. The sensor has a linear relationship between glucose concentration and response current in the range of 5–80 mM. The linear equation is $I(\mu A) = 7.268 + 0.507C$ (mM), the correlation coefficient $R^2 = 0.9911$, and the detection limit (LOD) is 0.537 μM (S/N = 3). As for fructose (**Figure 5B**), the detection range of the sensor is 2–20 mM, the linear equation is $I(\mu A) = 0.720 + 1.936C$ (mM), and the LOD is 1.630 μM. The sensor also showed similar current response to arabinose, galactose, mannose, and xylose (**Supplementary Figures S3–S6**). **Table 1** shows the analysis characteristics of the sensor to all sugars. In plants or agricultural products, the content of sugars is very high, ranging from a few millimoles to thousands of millimoles (Zhou et al., 2019). Therefore, for *in situ* detection of sugars in plants or agricultural products, the lower detection limit does not need to

be as low as micromolar level, while the upper detection limit needs to be as high as tens or even thousands of millimol level. **Table 2** shows the analytical performance of different enzyme-free glucose sensors reported previously. From **Table 2**, we can see that the highest upper detection limit of the developed enzyme-free glucose sensor is 19.6 mM (Jeong et al., 2018). The detection range for glucose of our sensor is 5–80 mM. Clearly, our sensor can be used for in situ detection of glucose in more plants and agricultural products. **Supplementary Table S1** shows the comparison of analytical performance of different enzyme-free fructose sensors. Our sensor shows the highest upper detection limit for fructose. **Supplementary Table S2** shows the analytical performance of enzyme-free sensors for arabinose, mannose, xylose, and galactose. The upper detection limit of our sensor (**Table 1**) for these four sugars is all higher than the corresponding reported sensors. Therefore, our sensor is also more suitable for in situ detection of fructose, arabinose, mannose, xylose, and galactose in agricultural products and plants.

In order to test the selectivity of the sensor, the prepared sensor was used to detect different interferences (20 mM), including malic acid, citric acid, tryptophan, leucine, lysine, magnesium chloride, sucrose, betaine, 3-indoleacetic acid, abscisic

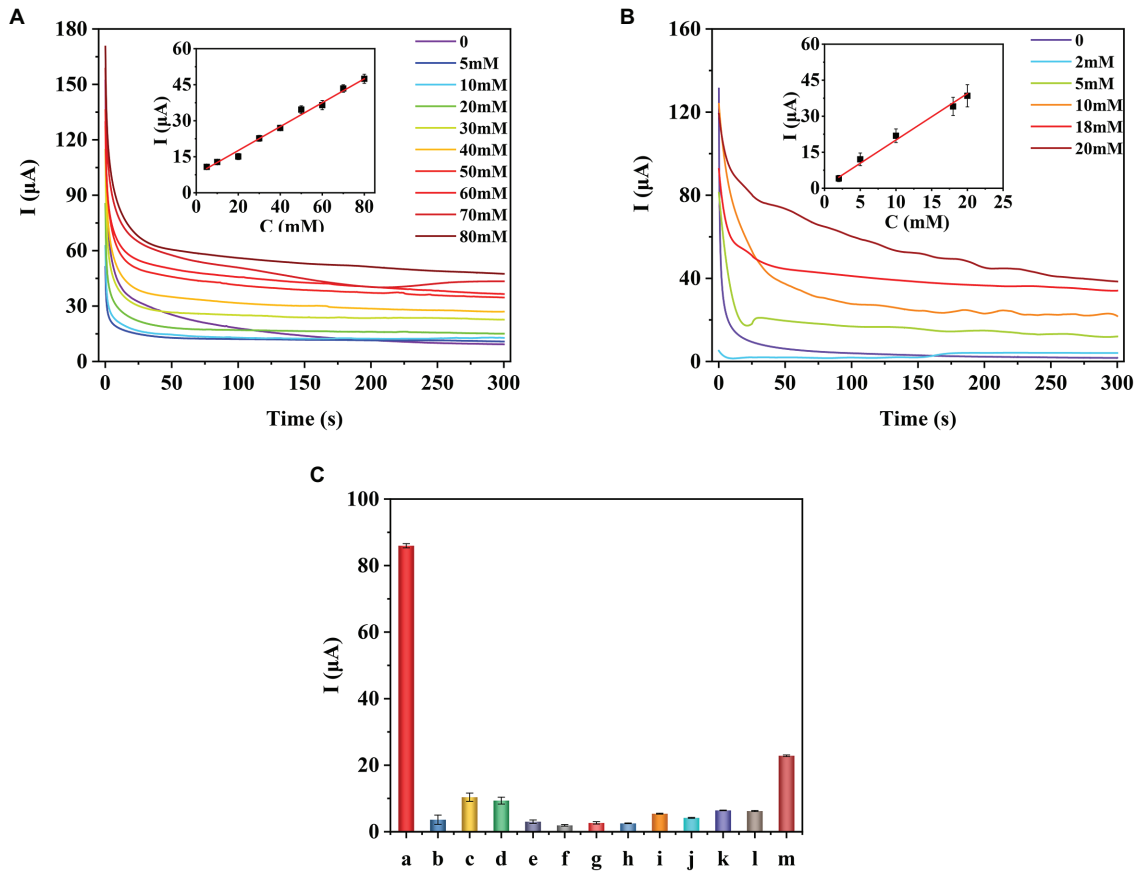


FIGURE 5 | I–t curves and calibration curves of the COOH-GR–COOH-MWNT–AuNPs/SPE sensor for the detection of different concentrations of glucose **(A)** and fructose **(B)**. Selectivity of the prepared the sugar sensor **(C)**. a-glucose, b-malic acid, c-citric acid, d-tryptophan, e-leucine, f-lysine, g-magnesium chloride, h-sucrose, i-Betaine, j-3-indoleacetic acid, k-abscisic acid, l-gibberellins, m-ascorbic acid.

TABLE 1 | Analytical characteristics of different sugars detected by the sugar sensor.

Sugar	Lineal range (mM)	Intercept	Slope	Calibration	R^2	LOD (mM)
Glucose	5–80	7.268 ± 0.635	0.507 ± 0.018	7.268 + 0.507C	0.9911	0.537
Fructose	2–20	0.720 ± 0.872	1.936 ± 0.112	0.720 + 1.936C	0.9901	1.630
Arabinose	2–50	3.905 ± 2.792	2.019 ± 0.086	3.905 + 2.019C	0.9910	1.811
Mannose	5–60	−13.812 ± 2.020	30.924 ± 1.542	−13.812 + 30.924lgC	0.9853	4.903
Xylose	2–40	1.270 ± 0.385	8.979 ± 0.390	1.270 + 8.979C	0.9888	0.693
Galactose	5–40	−16.943 ± 4.239	74.263 ± 3.360	−16.943 + 74.263lgC	0.9919	2.105

TABLE 2 | Comparison of analytical performance of different enzyme-free glucose sensors.

Electrode	Linear range (mM)	Detection limit (μM)	References
CNTs/AuNPs/GCE	0.002–19.6	0.5	Jeong et al., 2018
Cu/Ni/graphene/Ta	5×10^{-6}–2.174	0.0027	Cui et al., 2019
CuO/Nafion/GC	0.001–10	0.57	Pérez-Fernández et al., 2017
SPE/NiCo/C	5×10^{-4}–4.38	0.2	Wang et al., 2020
CuO/$Ni(OH)_2$/CC	0.05–8.50	0.31	Sun et al., 2020
MOF/CuO	0–6.535	0.15	Luo et al., 2020
Cu/Ni/Au	5×10^{-4}–3.0, 3.0–7.0	0.1	Liu et al., 2020
CuO NWs/GC	0.0125–4.29	4.17	Zhang, 2019
Au/CQDs	0.05–3	20	Han et al., 2020
COOH-GR-COOH-MWNT-AuNPs/SPE	5–80	540	This work

TABLE 3 | Recovery rate of glucose in apple juice ($n=3$).

Glucose initial (mM)	Added (mM)	Found (mM)	RSD (%)	Recovery (%)
12.075 (sensor)	10	22.092	3.45	100.17
10.800 (HPLC)	20	32.244	6.99	100.85
	30	41.294	5.69	97.40

Acid, gibberellins, and ascorbic acid. The results are shown in **Figure 5C**, and the current response of the sensor to glucose is significantly higher than other interfering substances, which proves that the sensor has good selectivity.

Under the same experimental conditions, the same SPE electrode was used to continuously measure the glucose solution of the same concentration (20 mM) for five times, and it can be seen (**Supplementary Figure S7A**) that the current response is relatively consistent (RSD = 7.36%). Five electrodes were used to measure the same concentration of glucose (**Supplementary Figure S7B**), and the RSD of the response current was 7.49%. These results show that the sensor has excellent reproducibility and high stability.

Detection in Real Samples

The standard addition method was adopted to detect glucose in apple juice. After 20 times of dilution, different concentrations of glucose were added. The results are shown in **Table 3**. The recovery results of other sugar are shown in **Supplementary Tables S3–S7**. The spiked recovery rate of glucose, fructose, arabinose, mannose, xylose, and galactose is 97.40–100.85%, 101.28–105.42%, 96.69–105.28%, 97.73–105.12%, 99.91–104.78%, and 98.42–104.89%, respectively, which show that our sensor has good practicality. Moreover, the results were also compared with those obtained by other methods. The initial concentration of glucose and fructose in the apple juice was also detected by the HPLC methods. The relative error of the results obtained by the as-prepared sensor and HPLC methods was 11.81 and 14.70%. The relative error between these two methods was smaller than 15%, which is considered acceptable (Artigues et al., 2021). For arabinose, mannose, xylose, and galactose, they cannot be separated by HPLC method. So the IC method was used to measure these sugars. As their content is very low, which has exceeded the detection range of IC, so the comparison result cannot be obtained. But the initial concentration of arabinose and mannose can be detected by our sensor, indicating that the as-prepared sensor has more potential in practical applications.

CONCLUSION

In summary, the developed enzyme-free reducing sugar sensor catalyzes the oxidation reaction of six sugars through the synergistic effect of graphene, carbon nanotubes, and gold nanoparticles. The detection range of the sensor for glucose, fructose, arabinose, mannose, xylose, and galactose is 5–80, 2–20, 2–50, 5–60, 2–40, and 5–40 mM, respectively. To our knowledge, the upper detection limit of our enzyme-free sugar sensor is the highest compared to previous studies, which is more suitable for in-situ detection of sugars in agricultural products and plants. This sensor is simple and portable and has good reproducibility and stability. Therefore, it will have broad practical application value in precision agriculture. With the introduction of various new nanomaterials, sensors with wider detection range, higher upper detection limit and better selection performance are expected to be developed, which is more suitable for the *in situ* quantification of sugars in plants and agricultural products.

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

KL and XW performed the experiments, analyzed the data, and wrote the original manuscript. BL, CW, PH, and HD helped to perform the experiments. AL and CZ supervised the project, designed the research, and wrote, reviewed, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The authors are thankful for the funding from the Key-Area Research and Development Program of Guang Dong Province (No. 2021B0707010002) and the National Natural Science Foundation of China (Grant No. 21974012).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.872190/full#supplementary-material

REFERENCES

Artigues, M., Gilabert-Porres, J., Texidó, R., Borrós, S., Abellà, J., and Colominas, S. (2021). Analytical parameters of a novel glucose biosensor based on grafted PFM as a covalent immobilization technique. *Sensors* 21:4185. doi: 10.3390/s21124185

Ayyub, O. B., Ibrahim, M. B., Briber, R. M., and Kofinas, P. (2013). Self-assembled block copolymer photonic crystal for selective fructose detection. *Biosens. Bioelectron.* 46, 124–129. doi: 10.1016/j.bios.2013.02.025

Biscay, J., Costa Rama, E., González García, M. B., Julio Reviejo, A., Pingarrón Carrazón, J. M., and García, A. C. (2012). Amperometric fructose sensor based on ferrocyanide modified screen-printed carbon electrode. *Talanta* 88, 432–438. doi: 10.1016/j.talanta.2011.11.013

Campos, A. M., Raymundo-Pereira, P. A., Mendonça, C. D., Calegaro, M. L., Machado, S. A. S., and Oliveira, O. N. Jr. (2018). Size control of carbon spherical shells for sensitive detection of paracetamol in sweat, saliva, and urine. *ACS Appl. Nano Mater.* 1, 654–661. doi: 10.1021/acsanm.7b00139

Cui, X., Lv, R., Sagar, R., Liu, C., and Zhang, Z. (2015). Reduced graphene oxide/carbon nanotube hybrid film as high performance negative electrode for supercapacitor. *Electrochim. Acta* 169, 342–350. doi: 10.1016/j.electacta.2015.04.074

Cui, D., Lin, S., Li, B., Li, M., Li, C., Xu, S., et al. (2019). Non-enzymatic glucose sensor based on micro-/nanostructured Cu/Ni deposited on graphene sheets. *J. Electroanal. Chem.* 838, 154–162. doi: 10.1016/j.jelechem.2019.03.005

de Sá, A. C., Cipri, A., González-Calabuig, A., Stradiotto, N. R., and Del Valle, M. (2016). Resolution of galactose, glucose, xylose and mannose in sugarcane bagasse employing a voltammetric electronic tongue formed by metals oxy-hydroxide/MWCNT modified electrodes. *Sensors Actuators B Chem.* 222, 645–653. doi: 10.1016/j.snb.2015.08.088

Gao, W., Emaminejad, S., Nyein, H. Y. Y., Challa, S., Chen, K., Peck, A., et al. (2016). Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature* 529, 509–514. doi: 10.1038/nature16521

Gizem Güneştekin, B., Medetalibeyoglu, H., Atar, N., and Lütfi Yola, M. (2020). Efficient direct-methanol fuel cell based on Graphene quantum dots/multi-walled carbon nanotubes composite. *Electroanalysis* 32, 1977–1982. doi: 10.1002/elan.202060074

Gota, T., Chowdhury, M., and Ojumu, T. (2017). Non-enzymatic fructose sensor based on Co3 O4 thin film. *Electroanalysis* 29, 2855–2862. doi: 10.1002/elan.201700503

Han, S., Gao, Y., Li, L., Lu, B., Zou, Y., Zhang, L., et al. (2020). Synergistic enhancement effects of carbon quantum dots and au nanoclusters for cathodic ECL and non-enzyme detections of glucose. *Electroanalysis* 32, 1155–1159. doi: 10.1002/elan.201900645

Hernaez, M. (2020). Applications of Graphene-based materials in sensors. *Sensors* 20:3196. doi: 10.3390/s20113196

Jeong, H., Nguyen, D. M., Lee, M. S., Kim, H. G., Ko, S. C., and Kwac, L. K. (2018). N-doped graphene-carbon nanotube hybrid networks attaching with gold nanoparticles for glucose non-enzymatic sensor. *Mater. Sci. Eng. C Mater. Biol. Appl.* 90, 38–45. doi: 10.1016/j.msec.2018.04.039

Kamran, U., Bhatti, H. N., Iqbal, M., Jamil, S., and Zahid, M. (2019a). Biogenic synthesis, characterization and investigation of photocatalytic and antimicrobial activity of manganese nanoparticles synthesized from Cinnamomum verum bark extract. *J. Mol. Struct.* 1179, 532–539. doi: 10.1016/j.molstruc.2018.11.006

Kamran, U., Heo, Y., Lee, J. W., and Park, S. (2019b). Functionalized carbon materials for electronic devices: a review. *Micromachines* 10:234. doi: 10.3390/mi10040234

Kamran, U., and Park, S. (2020). MnO2-decorated biochar composites of coconut shell and rice husk: An efficient lithium ions adsorption-desorption performance in aqueous media. *Chemosphere* 260:127500. doi: 10.1016/j.chemosphere.2020.127500

Kamran, U., and Park, S. (2021a). Acetic acid-mediated cellulose-based carbons: influence of activation conditions on textural features and carbon dioxide uptakes. *J. Colloid Interface Sci.* 594, 745–758. doi: 10.1016/j.jcis.2021.03.069

Kamran, U., and Park, S. (2021b). Chemically modified carbonaceous adsorbents for enhanced CO_2 capture: a review. *J. Clean. Prod.* 290:125776. doi: 10.1016/j.jclepro.2020.125776

Kamran, U., Bhatti, H. N., Noreen, S., Tahir, M. A., and Park, S. (2022). Chemically modified sugarcane bagasse-based biocomposites for efficient removal of acid red 1 dye: kinetics, isotherms, thermodynamics, and desorption studies. *Chemosphere* 291:132796. doi: 10.1016/j.chemosphere.2021.132796

Kamran, U., and Park, S. (2022). Hybrid biochar supported transition metal doped MnO2 composites: efficient contenders for lithium adsorption and recovery from aqueous solutions. *Desalination* 522:115387. doi: 10.1016/j.desal.2021.115387

Li, H., Yang, C., Zhu, X., and Zhang, H. (2017). A simple ratiometric fluorescent sensor for fructose based on complexation of 10-hydroxybenzo[h]quinoline with boronic acid. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 180, 199–203. doi: 10.1016/j.saa.2017.03.017

Li, L., and Sheen, J. (2016). Dynamic and diverse sugar signaling. *Curr. Opin. Plant Biol.* 33, 116–125. doi: 10.1016/j.pbi.2016.06.018

Liu, G., Zhao, J., Qin, L., Liu, S., Zhang, Q., and Li, J. (2020). Synthesis of an ordered nanoporous Cu/Ni/Au film for sensitive non-enzymatic glucose sensing. *RSC Adv.* 10, 12883–12890. doi: 10.1039/D0RA01224F

Luo, Y., Wang, Q., Li, J., Xu, F., Sun, L., Bu, Y., et al. (2020). Tunable hierarchical surfaces of CuO derived from metal-organic frameworks for non-enzymatic glucose sensing. *Inorg. Chem. Front.* 7, 1512–1525. doi: 10.1039/d0qi00104j

Najafianpour, N., and Dorranian, D. (2018). Properties of graphene/Au nanocomposite prepared by laser irradiation of the mixture of individual colloids. *Appl. Phys.* 124:805. doi: 10.1007/s00339-018-2236-7

Özcan, N., Karaman, C., Atar, N., Karaman, O., and Yola, M. L. (2020a). A novel molecularly imprinting biosensor including graphene quantum dots/multi-walled carbon nanotubes composite for interleukin-6 detection and electrochemical biosensor validation. *ECS J. Solid State Sci. Technol.* 9:121010. doi: 10.1149/2162-8777/abd149

Özcan, N., Medetalibeyoglu, H., Akyıldırım, O., Atar, N., and Yola, M. L. (2020b). Electrochemical detection of amyloid-β protein by delaminated titanium carbide MXene/multi-walled carbon nanotubes composite with molecularly imprinted polymer. *Mater. Today Commun.* 23:101097. doi: 10.1016/j.mtcomm.2020.101097

Pablos, J. L., Vallejos, S., Ibeas, S., Muñoz, A., Serna, F., García, F. C., et al. (2015). Acrylic polymers with pendant phenylboronic acid moieties as "turn-off" and "turn-On" fluorescence solid sensors for detection of dopamine, glucose, and fructose in water. *ACS Macro Lett.* 4, 979–983. doi: 10.1021/acsmacrolett.5b00465

Pérez-Fernández, B., Martin-Yerga, D., and Costa-García, A. (2017). Galvanostatic electrodeposition of copper nanoparticles on screen-printed carbon electrodes and their application for reducing sugars determination. *Talanta* 175, 108–113. doi: 10.1016/j.talanta.2017.07.026

Pohanka, M. (2020). Screen printed electrodes in biosensors and bioassays. A review. *Int. J. Electrochem. Sci.* 15, 11024–11035. doi: 10.20964/2020.11.19

Raymundo-Pereira, P. A., Campos, A. M., Mendona, C. D., Calegaro, M. L., and Oliveira, O. N. (2017). Printex 6L carbon nanoballs used in electrochemical sensors for simultaneous detection of emerging pollutants hydroquinone and paracetamol. *Sensors Actuators B Chem.* 252, 165–174. doi: 10.1016/j.snb.2017.05.121

Sharma, A., Arya, S., Chauhan, D., Solanki, P. R., and Khosla, A. (2020). Synthesis of Au-SnO2 nanoparticles for electrochemical determination of vitamin B12. *J. Mater. Res. Technol.* 9, 14321–14337. doi: 10.1016/j.jmrt.2020.10.024

Shekarchizadeh, H., Kadivar, M., and Ensafi, A. A. (2013). Rapid nonenzymatic monitoring of glucose and fructose using a CuO/multiwalled carbon nanotube nanocomposite-modified glassy carbon electrode. *Chin. J. Catal.* 34, 1208–1215. doi: 10.1016/s1872-2067(12)60586-5

Shi, K., and Shiu, K. K. (2002). Scanning tunneling microscopic and voltammetric studies of the surface structures of an electrochemically activated glassy carbon electrode. *Anal. Chem.* 74, 879–885. doi: 10.1021/ac010734

Shim, S. (2019). Acid–base chemistry of Porphyrin/Graphene oxide complex: role of electrostatic interaction. *Bull. Kor. Chem. Soc.* 40, 366–369. doi: 10.1002/bkcs.11700

Strakosas, X., Selberg, J., Pansodtee, P., Yonas, N., Manapongpun, P., Teodorescu, M., et al. (2019). A non-enzymatic glucose sensor enabled by bioelectronic pH control. *Sci. Rep.* 9:10844. doi: 10.1038/s41598-019-46302-9

Su, D., Cortie, M., and Wang, G. (2017). Fabrication of N-doped Graphene–Carbon nanotube hybrids from Prussian Blue for Lithium–Sulfur Batteries. *Adv. Energy Mater.* 7:1602014. doi: 10.1002/aenm.201602014

Sun, S., Shi, N., Liao, X., Zhang, B., Yin, G., Huang, Z., et al. (2020). Facile synthesis of CuO/Ni(OH)2 on carbon cloth for non-enzymatic glucose sensing. *Appl. Surf. Sci.* 529:147067. doi: 10.1016/j.apsusc.2020.147067

Tabatabaie, N., and Dorranian, D. (2016). Effect of fluence on carbon nanostructures produced by laser ablation in liquid nitrogen. *Appl. Phys.* 122:558. doi: 10.1007/s00339-016-0091-y

Tourani, S., Rashidi, A., Safekordi, A. A., Aghabozorg, H. R., and Khorasheh, F. (2015). Synthesis of reduced graphene oxide-carbon nanotubes (rGO-CNT) composite and its use as a novel catalyst support for hydro-purification of crude terephthalic acid. *Ind. Eng. Chem. Res.* 54, 7591–7603. doi: 10.1021/acs.iecr.5b01574

Wahjudi, P. N., Patterson, M. E., Lim, S., Yee, J. K., Mao, C. S., and Lee, W. N. P. (2010). Measurement of glucose and fructose in clinical samples using gas chromatography/mass spectrometry. *Clin. Biochem.* 43, 198–207. doi: 10.1016/j.clinbiochem.2009.08.028

Wang, C., Yuan, R., Chai, Y., Chen, S., Zhang, Y., Hu, F., et al. (2012). Non-covalent iron(III)-porphyrin functionalized multi-walled carbon nanotubes for the simultaneous determination of ascorbic acid, dopamine, uric acid and nitrite – ScienceDirect. *Electrochim. Acta* 62, 109–115. doi: 10.1016/j.electacta.2011.11.115

Wang, L., Hou, C., Yu, H., Zhang, Q., Li, Y., and Wang, H. (2020). Metal organic framework-derived nickel/cobalt-based nanohybrids for sensing non-enzymatic glucose. *ChemElectroChem* 7, 4446–4452. doi: 10.1002/celc.202001135

Wang, Q., Wang, T., Wang, J., Guo, W., Qian, Z., and Wei, T. (2018). Preparation of antistatic high-density polyethylene composites based on synergistic effect of graphene nanoplatelets and multi-walled carbon nanotubes. *Polym. Adv. Technol.* 29, 407–416. doi: 10.1002/pat.4129

Wu, A., Su, X., Fang, Y., Sun, J., and Chen, G. (2008). Electrogenerated chemiluminescence at bare glassy carbon electrode in basic media. *Electrochem. Commun.* 10, 1344–1346. doi: 10.1016/j.elecom.2008.05.011

Xu, D., Luo, L., Ding, Y., Jiang, L., Zhang, Y., Ouyang, X., et al. (2014a). A novel nonenzymatic fructose sensor based on electrospun LaMnO3 fibers. *J. Electroanal. Chem.* 727, 21–26. doi: 10.1016/j.jelechem.2014.05.010

Xu, F., Wang, F., Yang, D., Gao, Y., and Li, H. (2014b). Electrochemical sensing platform for L-CySH based on nearly uniform Au nanoparticles decorated graphene nanosheets. *Mater. Sci. Eng. C* 38, 292–298. doi: 10.1016/j.msec.2014.02.017

Xu, H., Xia, C., Wang, S., Han, F., Akbari, M. K., Hai, Z., et al. (2018). Electrochemical non-enzymatic glucose sensor based on hierarchical 3D Co3O4/Ni heterostructure electrode for pushing sensitivity boundary to a new limit. *Sensors Actuators B Chem.* 267, 93–103. doi: 10.1016/j.snb.2018.04.023

Yola, M. L., Atar, N., and Zcan, N. (2021). A novel electrochemical lung cancer biomarker cytokeratin 19 fragment antigen 21-1 immunosensor based on Si3N4/MoS2 incorporated MWCNTs and core–shell type magnetic nanoparticles. *Nanoscale* 13, 4660–4669. doi: 10.1039/d1nr00244a

Yola, M. L., and Atar, N. (2021). Novel voltammetric tumor necrosis factor-alpha (TNF-α) immunosensor based on gold nanoparticles involved in thiol-functionalized multi-walled carbon nanotubes and bimetallic Ni/Cu-MOFs. *Anal. Bioanal. Chem.* 413, 2481–2492. doi: 10.1007/s00216-021-03203-z

Zhang, D. (2019). Facile two-step electrodeposition synthesis of CuO nanowires for ultrasensitive non-enzymatic sensing of glucose. *Int. J. Electrochem. Sci.* 14, 10835–10847. doi: 10.20964/2019.12.26

Zhou, Y., Xu, K., and Wang, Q. (2019). Determination of contents of glucose, fructose, sucrose, and sorbitol in vegetables by UPLC-MS/MS. *Chem. Bioeng.* 36, 66–68. doi: 10.3969/j.issn.1672-5425.2019.03.014

Zhang, K., Li, G., Feng, L., Wang, N., Guo, J., Sun, K., et al. (2017). Ultralow percolation threshold and enhanced electromagnetic interference shielding in poly(L-lactide)/multi-walled carbon nanotube nanocomposites with electrically conductive segregated networks. *J. Mater. Chem. C Mater. Opt. Electr. Devices* 5, 9359–9369. doi: 10.1039/c7tc02948a

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 May 2022
doi: 10.3389/fpls.2022.849821

A Shape Reconstruction and Measurement Method for Spherical Hedges Using Binocular Vision

*Yawei Zhang[1], Jin Gu[1], Tao Rao[1], Hanrong Lai[1], Bin Zhang[1], Jianfei Zhang[4] and Yanxin Yin[2,3]**

[1] College of Engineering, China Agricultural University, Beijing, China, [2] Research Center of Intelligent Equipment, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, [3] National Research Center of Intelligent Equipment for Agriculture, Beijing, China, [4] Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing, China

The center coordinate and radius of the spherical hedges are the basic phenotypic features for automatic pruning. A binocular vision-based shape reconstruction and measurement system for front-end vision information gaining are built in this paper. Parallel binocular cameras are used as the detectors. The 2D coordinate sequence of target spherical hedges is obtained by region segmentation and object extraction process. Then, a stereo correcting algorithm is conducted to keep two cameras to be parallel. Also, an improved semi-global block matching (SGBM) algorithm is studied to get a disparity map. According to the disparity map and parallel structure of the binocular vision system, the 3D point cloud of the target is obtained. Based on this, the center coordinate and radius of the spherical hedges can be measured. Laboratory and outdoor tests on shape reconstruction and measurement are conducted. In the detection range of 2,000–2,600 mm, laboratory test shows that the average error and average relative error of standard spherical hedges radius are 1.58 mm and 0.53%, respectively; the average location deviation of the center coordinate of spherical hedges is 15.92 mm. The outdoor test shows that the average error and average relative error of spherical hedges radius by the proposed system are 4.02 mm and 0.44%, respectively; the average location deviation of the center coordinate of spherical hedges is 18.29 mm. This study provides important technical support for phenotypic feature detection in the study of automatic trimming.

Keywords: spherical hedges, shape reconstruction, binocular vision, dimension measurement, 3D point cloud

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Paul Barry Hibbard,
University of Essex, United Kingdom
Lihui Wang,
China Conservatory, China
Sixun Chen,
Hokkaido University, Japan

***Correspondence:**
Yanxin Yin
yinyx @nercita.org.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *06 January 2022*
Accepted: *14 March 2022*
Published: *04 May 2022*

Citation:
Zhang Y, Gu J, Rao T, Lai H, Zhang B, Zhang J and Yin Y (2022) A Shape Reconstruction and Measurement Method for Spherical Hedges Using Binocular Vision. Front. Plant Sci. 13:849821. doi: 10.3389/fpls.2022.849821

INTRODUCTION

With the vigorous development of urban greening, trimming or pruning hedges to desired shape regulars is one of the major tasks in urban plant landscape construction. Manual trimming using large scissors or power tools causes a significant load on the person executing this task. The semi-automated trimmer, however, also needs a driver operating, consumes most time, and is difficult to control working accuracy. Therefore, the development of automatic and intelligent pruning robots has drawn increasing attention.

To automatically trim hedges, finding the basic phenotypic information of hedges is the key. In a complex outdoor environment, an adaptive hedge horizontal cross-section center

detection algorithm was proposed to obtain the hedge's horizontal cross-section center in real time by inputting the top view image of the hedge. This detection algorithm could be truly applied in the vehicle-mounted system (Li et al., 2022). A TrimBot2020 robotic platform equipped with a pentagon-shaped rig of five pairs of stereo cameras was developed for navigation and 3D reconstruction, which can build the model of bush or hedges and be used as the input for the trimming operation (Strisciuglio et al., 2018). An arm-mounted vision approach was studied to scan a specified shape and fit it into the reconstructed point cloud, and then, a co-mounted trimming tool could cut the bush using an automatically planned trajectory, which ensured flexibility *via* a vision-based shape fitting module that allows fitting an arbitrary mesh into a bush at hand (Kaljaca et al., 2019a,b). Besides, the binocular vision system has great application in picking robots for object recognition and orientation. A litchi-picking robot based on binocular vision was developed to identify and locate the target and then provide information for collision-free motion planning. The results show that the success rate of path determination is 100% for the laboratory's picking scene (Ye et al., 2021). Herein, vision sensing technology was widely used in characteristic recognition of fruits and vegetables and movement navigation of picking robots, such as tomatoes, apples, and Hangzhou White Chrysanthemums (Ji et al., 2017; Lili et al., 2017; Yang et al., 2018; Jin et al., 2020). From the above research, it can be concluded that binocular stereo vision technology has been widely used in agricultural robotics for three-dimensional (3D) reconstruction, measurement, navigation, etc. As the "eye" of the pruning robot, the shape reconstruction and dimension measurement of target objects provide a crucial information for the follow-up operation.

In this paper, a parallel binocular vision is constructed to complete the 3D reconstruction of spherical hedges, and high accuracy is achieved in both spherical center positioning and radius measurement. The 3D reconstruction contains two-dimensional (2D) image extraction, binocular camera calibration, stereo correcting, stereo matching, and sharp reconstruction. Herein, in this paper, stereo matching is a key technology of shape reconstruction, and an improved semi-global block matching (SGBM) algorithm was proposed in this study to get a good disparity map. Based on this, the center coordinate of spherical hedges and their radius is finally realized by processing the point cloud data.

MATERIALS AND METHODS

Description of the Measurement System

To obtain point cloud information and reshape spherical hedges, a binocular vision system is used for measurement. The binocular vision system consists of two RMONCAM G200 cameras and a supporting platform. The cameras are mounted on the slider, and the positions of the cameras can be moved on the slider rail. The distance between two cameras can be set to 80, 100, 120, 140, and 160 mm. All experiments are involved in this paper, and the distance between the two cameras is set to 140 mm. The shape reconstruction and measurement system are programmed using Microsoft Visual Studio 2015, OpenCV3.4.10, and MATLAB2018a. The focus length, maximum frame rate, pixel size, and image resolution of a utilized camera are 2.8 mm, 60 fps, 3.0 μm × 3.0 μm, and 1,920 × 1,200 pixels, respectively. **Figure 1** shows the schematic diagram of the binocular vision system.

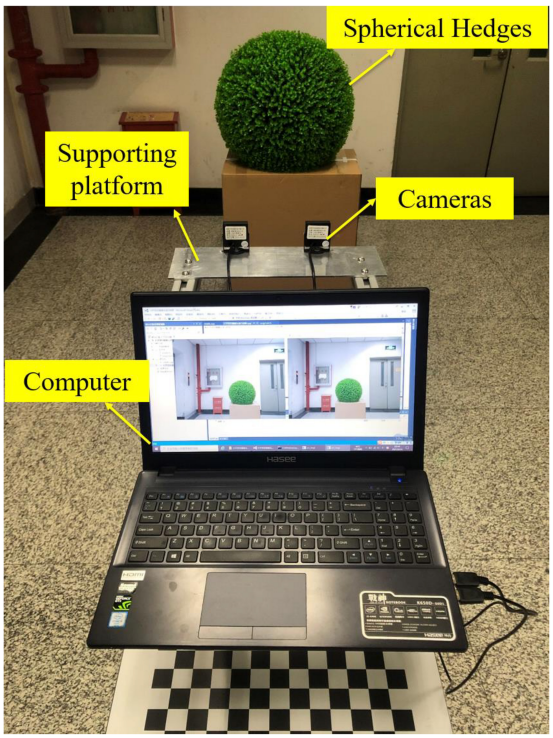


FIGURE 1 | Schematic diagram of the binocular vision system.

When conducting experiments, the spherical hedges are placed in front of the cameras. Then, the system captures the current images. Next, the images are transmitted to the computer. Afterward, image processing is called to obtain the point cloud data of spherical hedges. Based on this, the shape reconstruction graph is obtained. Finally, the radius and center coordinate of spherical hedges are calculated. **Figure 2** shows the flowchart of the measurement system.

Camera Calibration and Image Processing

Monocular Vision Calibration

Camera calibration is an important task because it directly determines the accuracy of 3D reconstruction (Long and Dongri, 2019). According to Zhang's camera plane calibration method, the calibration test of a monocular camera is carried out first. **Figure 3** presents the schematic diagram of pinhole imaging, $O_C - X_C Y_C Z_C$ is the camera coordinate system and $O_W - X_W Y_W Z_W$ is the world coordinate system; $O_1 - UV$is the pixel coordinate system and $O_2 - XY$ is the image coordinate system. $P(x_w, y_w, z_w)$ is the world coordinate of point P, and its

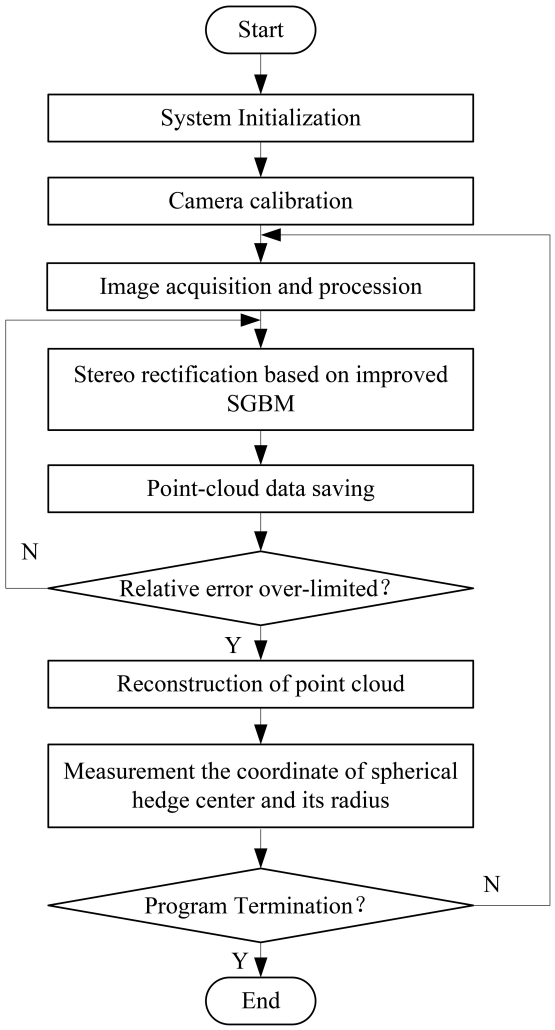


FIGURE 2 | Flow chart of the measurement system.

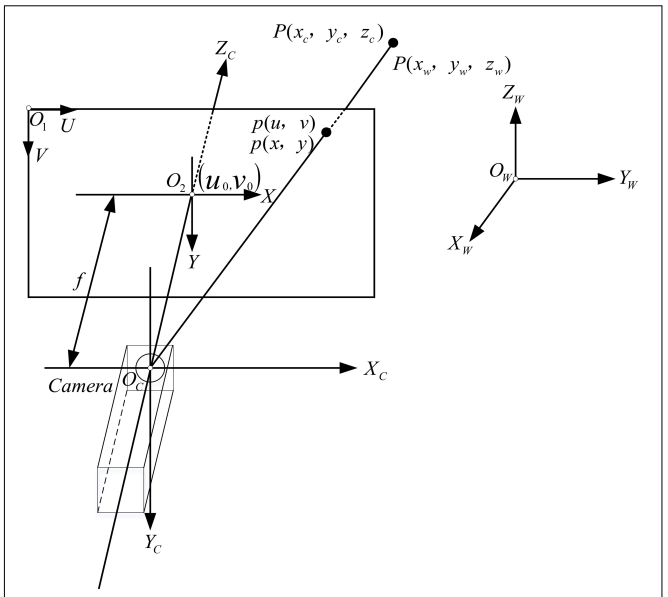


FIGURE 3 | Schematic diagram of pinhole imaging.

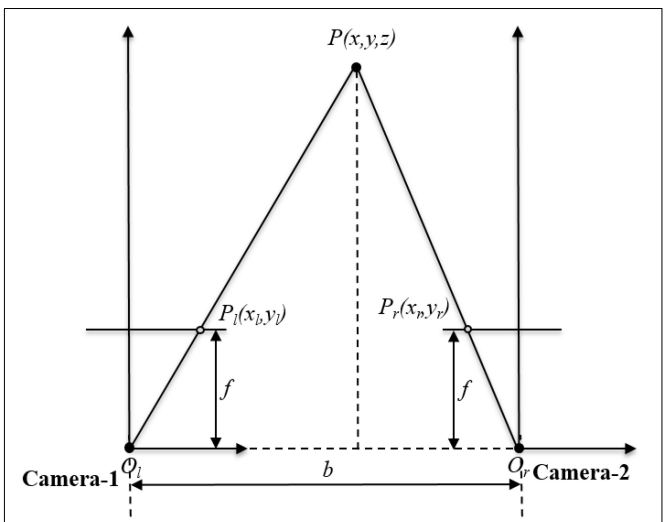


FIGURE 4 | The parallel structure of the binocular vision system.

corresponding camera coordinate in camera is $P(x_c, y_c, z_c)$ and its pixel coordinate is $p(u, v)$.

Converting from world coordinate system to pixel coordinate system needs to follow several transformations: transformation between world coordinate system and camera coordinate system; transformation between camera coordinate system and image coordinate system; and transformation between image coordinate system and pixel coordinate system.

The transformation between pixel coordinate system and image coordinate system is expressed as

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$

where s_x is the pixel size of 1 mm in the x-direction of $O_2 - XY$ and s_y is the pixel size of 1 mm in the y-direction of $O_2 - XY$.

The transformation between the camera coordinate system and image coordinate can be obtained from the pinhole imaging theory. It is formularized as

$$z_c \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} \tag{2}$$

where f is the focal length of the camera.

The transformation between the camera coordinate system and the world coordinate system can be obtained through

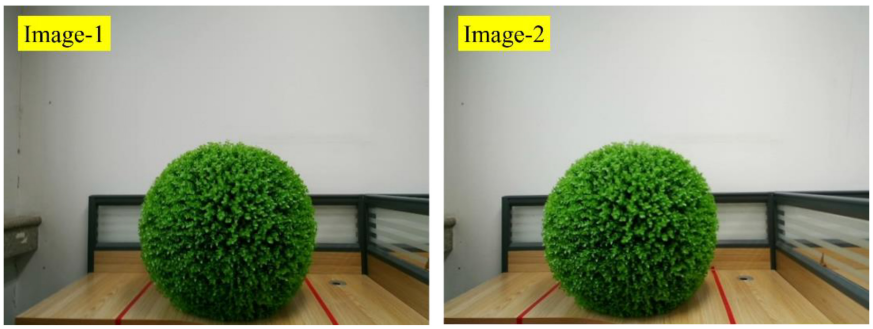


FIGURE 5 | The Image-1 and Image-2.

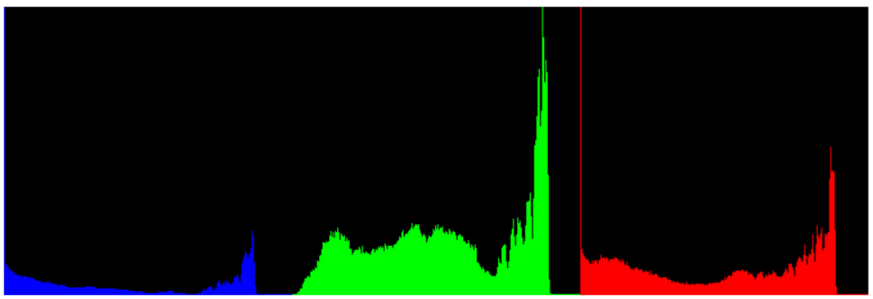


FIGURE 6 | RGB color histogram of Image-2.

rotation and translation. The transformation relationships are expressed as

$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & T \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \tag{3}$$

where $R$and T represent the rotation matrix and the horizontal movable matrix.

Herein, the transformation between world coordinate system to pixel coordinate system can be determined by

$$z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = M_1 M_2 \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

$$= M \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \tag{4}$$

where $M_1 = \begin{pmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, $M_2 = \begin{pmatrix} R & T \\ 0^T & 1 \end{pmatrix}$, $M = M_1 \bullet M_2$, $f_x = f \bullet s_x$, $f_y = f \bullet s_y$. The f_x , f_y , u_0 , and v_0 are camera intrinsic parameters, and thus, M_1 represents the camera's intrinsic parameter matrix. The M_2 represents the camera's extrinsic parameter matrix; hence, M represents the projection matrix of the camera.

Moreover, a high-order polynomial model is adopted to correct the image distortion. The high-order polynomial model is expressed as

$$\begin{bmatrix} x_c - x_0 \\ y_c - y_0 \end{bmatrix} = L(r) \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \tag{5}$$

where $L(r) = 1 + k_1 r + k_2 r^2 + k_3 r^3 + \ldots$, $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$, x and y refer to the horizontal and vertical coordinate values before correction, respectively, x_c and y_c refer to the horizontal and vertical coordinate values after correction, respectively, x_0 and y_0 refer to coordinate values of the center of the distorted image. Herein, a polynomial distortion correction model of the camera can be expressed as

$$\begin{cases} x_c = x\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \ldots\right) + 2p_1 xy + p_2\left(r^2 + 2x^2\right) \\ y_c = y\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \ldots\right) + 2p_2 xy + p_1\left(r^2 + 2y^2\right) \end{cases} \tag{6}$$

where k_1, k_2, and k_3 are radial distortion coefficients, p_1 and p_2 are tangential distortion coefficients. Herein, k_1, k_2, k_3, p_1, and p_2 are also camera intrinsic parameters.

Herein, the camera calibration toolbox (Toolbox_Calib) in MATLAB is used for monocular vision calibration. The calibration process of a monocular vision camera is as follows: image calibration, calibration chessboard extraction, corner

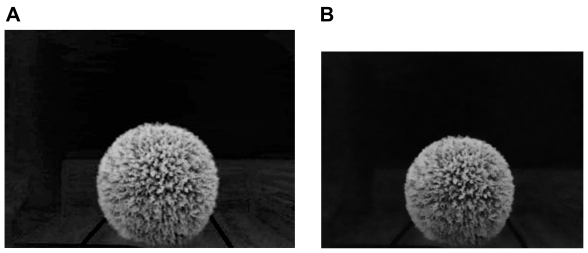


FIGURE 7 | (A) The gray image of Image-2 obtained after the Ultra-green algorithm. **(B)** The bilateral filtered image of Image-2.

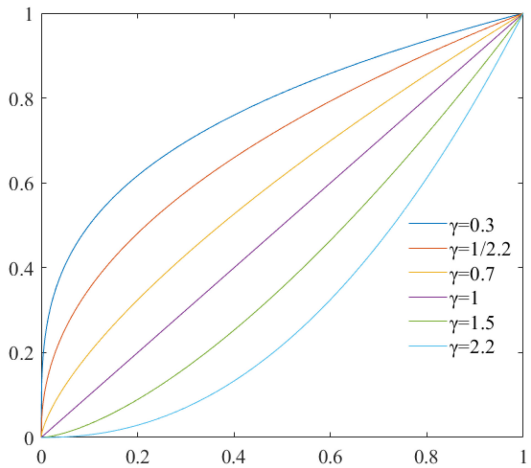


FIGURE 8 | The gamma transform graph at different γ values.

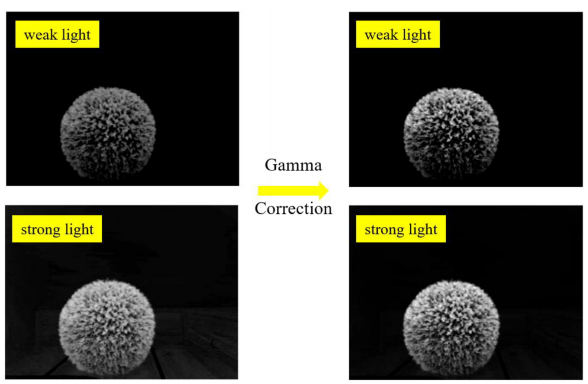


FIGURE 9 | The enhanced contrast results of spherical hedge images in weak light and strong light.

points extraction, intrinsic and extrinsic parameters calculation, and calibration error analysis.

Binocular Vision Calibration

The binocular vision calibration is conducted based on the monocular vision calibration; through calibration test, the intrinsic matrix and extrinsic matrix of a camera can be obtained.

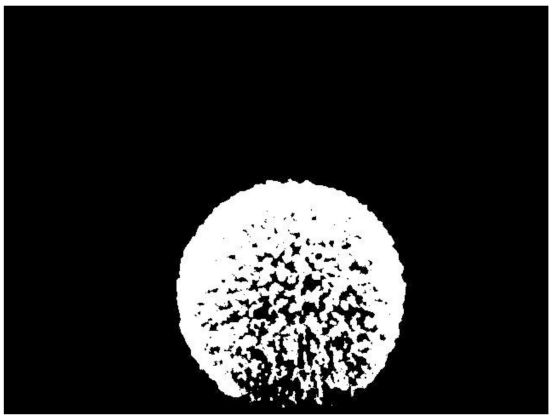


FIGURE 10 | The Image-2 after binarization.

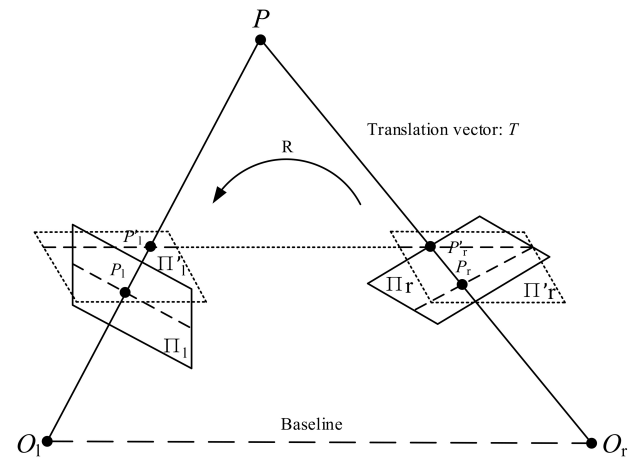


FIGURE 11 | Schematic diagram of Bouguet's algorithm.

In this paper, a parallel binocular stereo vision system is built. The two cameras are the same and mounted at the same height, and its front end is parallel and level. The parallel structure of the binocular vision system is shown in **Figure 4**. The left camera is called the Camera-1 and the left camera is called the Camera-2. Set the camera coordinate of Camera-1 as the reference world coordinate system. As indicated above, $P(x_w, y_w, z_w)$is the world coordinate of point P. Its corresponding image coordinate in Camera-1 is $p_l(x_l, y_l)$ and its corresponding image coordinate in Camera-2 is $p_r(x_r, y_r)$.

According to the principle of similar triangles, it can be obtained as

$$\begin{Bmatrix} x_w \\ y_w \\ z_w \end{Bmatrix} = \begin{Bmatrix} \frac{z}{f}x_l \\ \frac{z}{f}y_l \\ \frac{bf}{x_l - x_r} \end{Bmatrix} \tag{7}$$

where b is the baseline distance of Camera-1 and Camera-2, f is the focal length of the camera, and $x_l - x_r$ is the disparity value.

As in Equation 3, the transformation between the camera coordinate system of Camera-1 and world coordinate can be

obtained through rotating vector R_l and translation vector T_l, and the transformation between camera coordinate system of Camera-2 and world coordinate can be obtained through rotating vector R_r and translation vector T_r. Therefore, the transformation between camera coordinate systems of Camera-1 and Camera-2 can be represented as

$$\begin{cases} R = R_l R_r^{-1} \\ T = T_l - R_l R_r^{-1} T_r \end{cases} \tag{8}$$

The pixel coordinates of point P in Camera-1 and Camera-2 are $p_l(u_l, v_l)$ and $p_r(u_r, v_r)$, respectively. According to Equation 4, the transformation between world coordinate system to pixel coordinate system can be represented as

$$z_{cl}\begin{pmatrix} u_l \\ v_l \\ 1 \end{pmatrix} = \begin{pmatrix} f_{xl} & 0 & u_{0l} & 0 \\ 0 & f_{yl} & v_{0l} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} R_l & T_l \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}^1 & a_{12}^1 & a_{13}^1 & a_{14}^1 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 & a_{24}^1 \\ a_{31}^1 & a_{32}^1 & a_{33}^1 & a_{34}^1 \end{pmatrix}\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \tag{9}$$

$$z_{cr}\begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix} = \begin{pmatrix} f_{xr} & 0 & u_{0r} & 0 \\ 0 & f_{yr} & v_{0r} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} R_r & T_r \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}^2 & a_{12}^2 & a_{13}^2 & a_{14}^2 \\ a_{21}^2 & a_{22}^2 & a_{23}^2 & a_{24}^2 \\ a_{31}^2 & a_{32}^2 & a_{33}^2 & a_{34}^2 \end{pmatrix}\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \tag{10}$$

To solve the world coordinate $\left[x_w, y_w, z_w\right]^T$ of point P, taking the optical central position of Camera-1 as origin, an inhomogeneous linear equation is obtained through getting rid of Z_{cl} and Z_{cr} in Equations 9, 10.

$$\begin{cases} \left(u_l a_{31}^1 - a_{11}^1\right) x_W + \left(u_l a_{32}^1 - a_{12}^1\right) y_W + \left(u_l a_{33}^1 - a_{13}^1\right) z_W = a_{14}^1 - u_l a_{34}^1 \\ \left(v_l a_{31}^1 - a_{21}^1\right) x_W + \left(v_l a_{32}^1 - a_{22}^1\right) y_W + \left(v_l a_{33}^1 - a_{23}^1\right) z_W = a_{24}^1 - v_l a_{34}^1 \\ \left(u_{\mathrm{r}} a_{31}^2 - a_{11}^2\right) x_W + \left(u_r a_{32}^2 - a_{12}^2\right) y_W + \left(u_r a_{33}^2 - a_{13}^2\right) z_W = a_{14}^2 - u_r a_{34}^2 \\ \left(v_{\mathrm{r}} a_{31}^2 - a_{21}^2\right) x_W + \left(v_r a_{32}^2 - a_{22}^2\right) y_W + \left(v_r a_{33}^2 - a_{23}^2\right) z_W = a_{24}^2 - v_r a_{34}^2 \end{cases} \tag{11}$$

Up to now, for one point in space, as long as we obtain its pixel coordinates in Camera-1 and Camera-2, its world coordinates can be solved by Equation 11.

Region Segmentation and Object Extraction

After calibration, the binocular vision system can be used to capture images. The images captured by Camera-1 and Camera-2 are called Image-1 and Image-2, respectively. **Figure 5** shows Image-1 and Image-2.

Take Image-2 as an example to introduce the hedges extraction process. The RGB color histogram of Image-2 is shown in **Figure 6**, which shows that green color accounts for the largest proportion. Ultra-green extraction of green plant images has a good effect on distinguishing the green plants from the surrounding environment, and it is the most commonly used grayscale method for crop recognition or weed recognition. The excess green index (*ExG*) of ultra-green algorithm is set to $E\,\mathrm{x}G = 2G - R - B$.

Figure 7A is the 2G-R-B gray image of Image-2. The bilateral filtering for image denoising is used for image noise removal. **Figure 7B** is the bilateral filtered image of Image-2, which shows that the image boundary features can be most reserved.

Then, gamma correction was studied to enhance the contrast between the target hedges and the surrounding environment under strong light and weak light. The gamma formula can be expressed as

$$y = \left(x + esp\right)^{\gamma} \tag{12}$$

where, $x \in [0, 1]$, $y \in [0, 1]$, *esp* is the compensation factor, and γ is the gamma coefficient.

Figure 8 shows the grayscale mapping relationship between the output image and the input image with different γ values.

From **Figure 8**, it can be seen that different γ values should be used when performing gamma transformations for images with different grayscale distributions. In this paper, the contrast has been enhanced to some extent after gamma correction as shown in **Figure 9**, when γ equals 1.5.

At last, the image binarization best treatment threshold is obtained using the maximum between-class variance method (OTSU), hereafter, the 2D coordinate sequence of spherical hedges can be obtained from Image-2 (Caraffa et al., 2015). **Figure 10** shows the binary image of Image-2.

Shape Reconstruction and Measurement

Stereo Image Rectification

It is difficult to align the two cameras in this binocular vision system to be perfectly parallel (Wu et al., 2017). After binocular vision calibration, the stereo image rectification is used based on Bouguet's algorithm to ensure that the cameras are completely parallel. **Figure 11** shows the algorithmic principles of Bouguet's algorithm. The plane Π_l and plane Π_r are the image planes of Camera-1 and Camera-2 before polar correction, and the plane $\Pi_l^{'}$ and plane $\Pi_r^{'}$ are the image planes of Camera-1 and Camera-2 after polar correction. The $p_l^{'}$ and $p_r^{'}$ are the pixel coordinates of point P in the plane $\Pi_l^{'}$ and plane$\Pi_r^{'}$. The rotating vector R and translation vector T of the camera coordinate systems of Camera-1 and Camera-2 are obtained from camera calibration results.

In **Figure 11**, the practical binocular vision system can be corrected to a parallel binocular parallel system by multiplying the coordinate systems of Camera-1 and Camera-2 with their respective stereo correction matrices (R_{rect}) as follows

$$\begin{cases} R_l^{'} = R_{rect} \cdot R_l \\ R_r^{'} = R_{rect} \cdot R_r \end{cases} \tag{13}$$

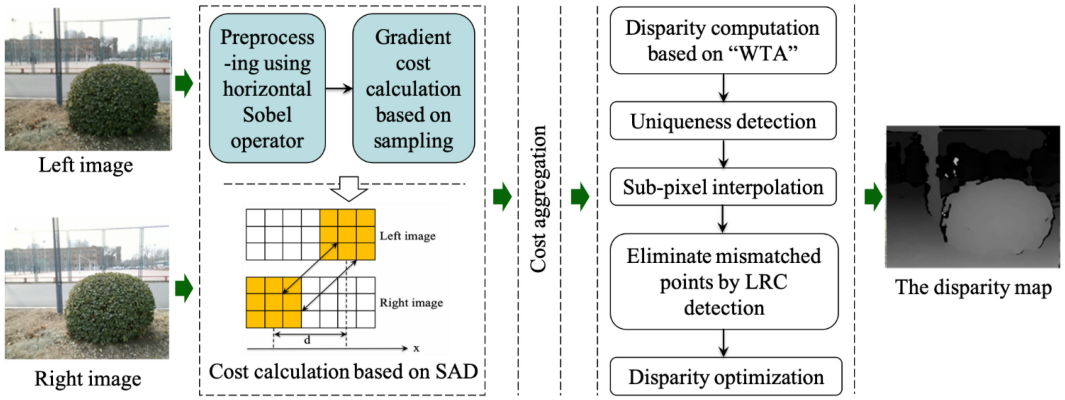


FIGURE 12 | Flow chart of the improved SGBM algorithm.

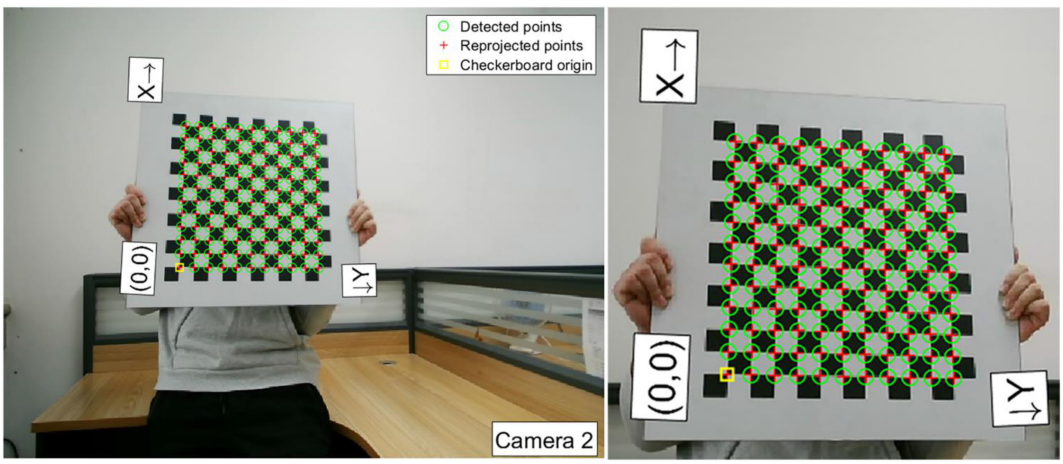


FIGURE 13 | Corner extraction results.

where $\begin{cases} R_l = R^{1/2} \\ R_r = R^{-1/2} \end{cases}$, $R_{rect} = \begin{bmatrix} \left(\frac{T}{||T||}\right)^T \\ \left(\frac{T}{||T||} \times \frac{[-T_y, T_x, 0]}{\sqrt{T_x^2+T_y^2}}\right)^T \end{bmatrix}$, $T = \left[T_x, T_y, T_z\right]^T$.

Shape Reconstruction

According to the morphological characteristics of spherical hedges, the surface fitting model is established by the SGBM algorithm. The SGBM algorithm is a classic semi-global matching algorithm, and this method has the advantages of both stereo matching quality and processing rates.

In the study of Romaniuk and Roszkowski (2014), the energy function of the SGBM algorithm can be represented as

$$E(D) = \sum_P (C(p, D_p)) + \sum_{q \in N_p} P_1 I\left[\left|D_p - D_q\right| = 1\right] + \sum_{q \in N_p} P_2 I\left[\left|D_p - D_q\right| > 1\right] \quad (14)$$

where $C(p, D_p)$ indicates matching cost value, N_p indicates pixels adjacent to point P, and P_1 and P_2 are penalty coefficient.

Considering operating efficiency, N_p is set to 8. The 2D search problem is divided into eight one-dimensional problems, thus using dynamic programming to treat each one-dimensional problem separately. When disparity is d, the matching cost value of point P in the r direction can be represented as

$$\begin{aligned} L_r(p, d) = C(p, d) + \min(&L_r(p - r, d), L_r(p - r, d - 1) \\ &+P_1, L_r(p - r, d + 1) + P_1, \min_i L_r(p - r, i) + P_2) \\ &- \min_k L_r(p - r, k) \end{aligned} \quad (15)$$

where $C(p, d)$ is the matching cost value when disparity is equal to d, $\min(L_r(p - r, d), L_r(p - r, d - 1) + P_1, L_r(p - r, d + 1)$ indicates the minimum matching cost value of previous

TABLE 1 | The intrinsic parameters and distortion coefficients of Camera-1 and Camera-2.

No.	f_x (pixel)	f_y (pixel)	u_0 (pixel)	v_0 (pixel)	k_1	k_2	p_1	p_2
Camera-1	510.2063	510.0911	331.6530	243.6460	0.0678	−0.0615	0.0017	0.0004103
Camera-2	505.7195	505.9684	329.6185	248.7792	0.0884	−0.0980	0.002834	0.0007345

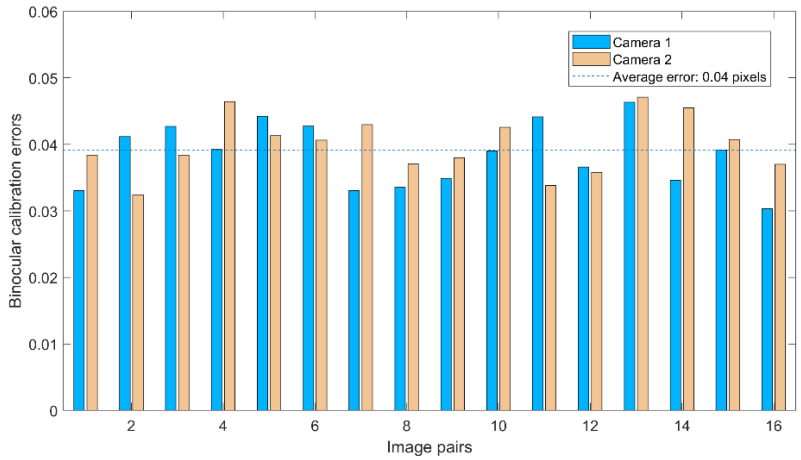


FIGURE 14 | Binocular calibration errors of each image pairs.

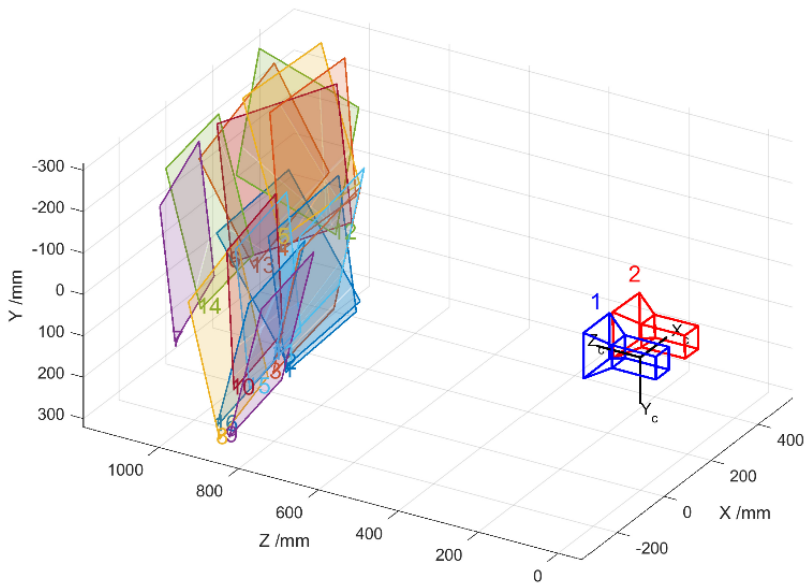


FIGURE 15 | The position and attitude relationship between cameras and calibration chessboard.

matching point pixel of point P in r direction, and $P_1, \min_i L_r(p-r,i) + P_2) - \min_k L_r(p-r,k)$ is the constraint.

Then, the matching cost values on each path were calculated and the total sum according to the SGBM algorithm was taken. The sum of matching cost value can be expressed by

$$S(p,d) = \sum_r L_r(p,d) \tag{16}$$

In the study of Hong and Ahn (2020), the optimal disparity d is corresponding to the minimum sum of matching cost value.

This study improves the SGBM algorithm by the following two main areas: occlusion detection and disparity optimization. The left-right consistency (LRC) method is used to remove the mismatch points, and the bilateral filtering algorithm is used to fill the holes in the disparity map. Then, the corresponding point cloud coordinates of the parallax map are calculated. **Figure 12** shows the flowchart of the improved SGBM algorithm.

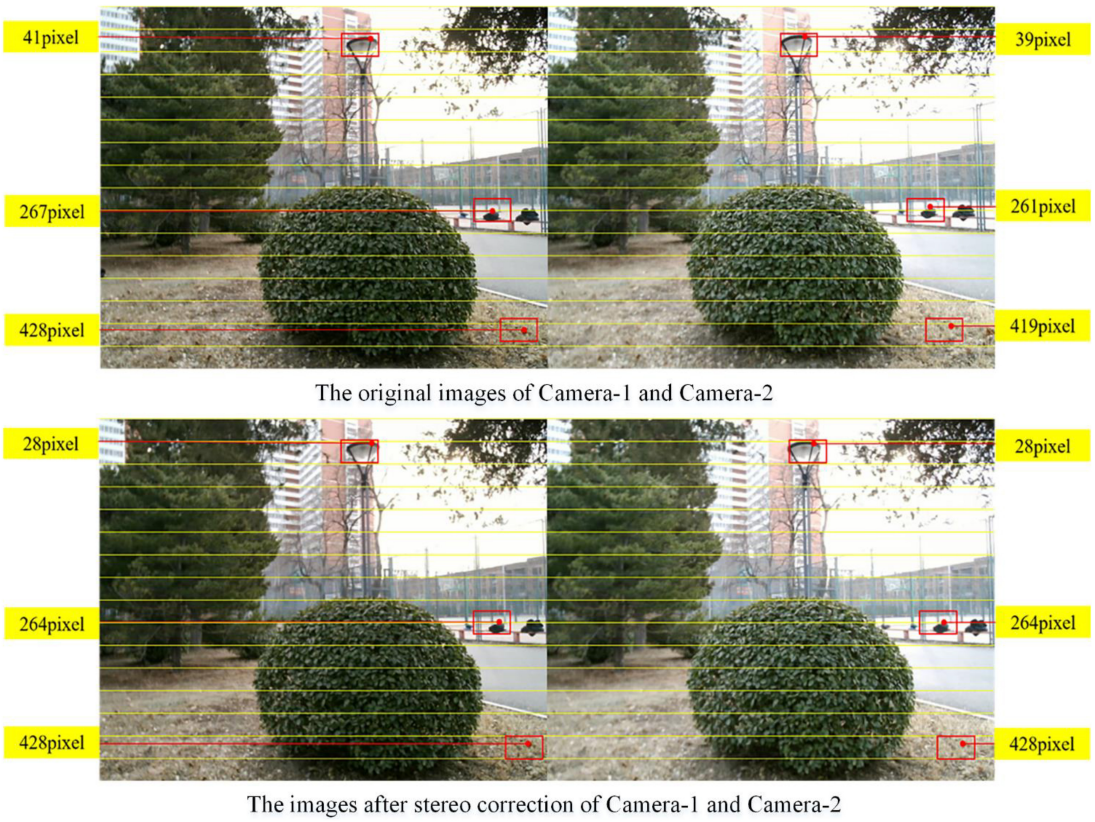


FIGURE 16 | Comparison of before and after stereo correction.

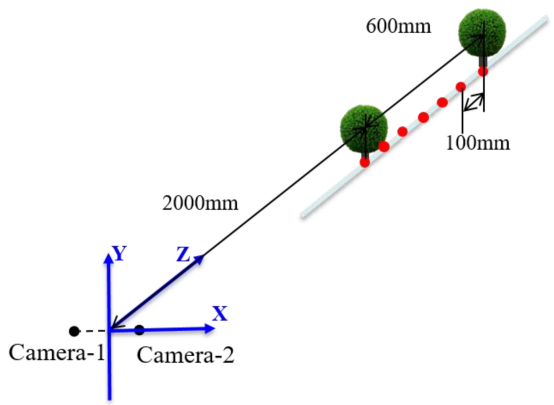


FIGURE 17 | Schematic diagram of laboratory test.

Occlusion detection based on the LRC is used to detect the disparity of all pixels in an image. When the disparity in the left and right imaging planes is inconformity, the pixels are regarded as the occluded points. To figure out occluded points, the disparity error is defined as

$$d(q) \neq -d(q + d(q)) \tag{17}$$

where $d(q)$ is the disparity of pixel q in the left imaging plane (Camera-1), $d(q + d(q))$ is the disparity of the corresponding pixel in the right imaging plane (Camera-2) when the disparity of pixel q is d.

Disparity optimization refers to filling the holes in the disparity map. After the occlusion detection, mismatch points or occluded points are removed, and thus, some pixels have no disparity value. Meanwhile, the depth of occluded points removed by the LRC detection is greater than the depth of the object that occludes it. Therefore, the disparity of occluded points can be estimated according to the non-occluded pixels and then fill it to the disparity map. Since disparity map-based hole filling is easily led to creating stripes, an edge keeping filter is used to reduce noise and save edge information of image well. The disparity processed by bilateral filtering can be expressed as

$$I_p^{bf} = \frac{\sum_{q \in S} G_{\sigma_s}(||\text{p-q}||) G_{\sigma_r}(|I_p - I_q|) I_q}{W_p^{bf}} \tag{18}$$

where, σ_s and σ_r are smooth parameters in the spatial domain and pixel range, I_p and I_q are input disparities of pixel p and pixel q, and W_p^{bf} is the bilateral filtering weight.

TABLE 2 | Laboratory test results of center coordinate and its radius.

No.	AR (mm)	MR (mm)	AER (mm)	RER (%)	AL (mm)	ML (mm)	LD (mm)
1	300	301.10	1.10	0.37	(100, −300, 2,000)	(101.54, −299.03, 2,018.79)	18.88
2	300	301.24	1.24	0.41	(100, −300, 2,100)	(102.17, −302.5, 2,111.72)	12.17
3	300	301.71	1.71	0.57	(100, −300, 2,200)	(103.3, −303.7, 2,211.05)	12.03
4	300	302.24	2.24	0.75	(100, −300, 2,300)	(103.25, −305.21, 2,318.11)	19.12
5	300	300.67	0.67	0.22	(100, −300, 2,400)	(104.03, −305.67, 2,414.41)	16.00
6	300	301.08	1.08	0.36	(100, −300, 2,500)	(104.36, −306.59, 2,514.02)	16.10
7	300	297.00	3.00	1.00	(100, −300, 2,600)	(105.27, −308.54, 2,613.91)	17.15

AR, actual radius; MR, measured radius; AER, absolute error of radius; RER, relative error of radius; AL, actual location; ML, measured location; LD, location deviation. The average absolute error of radius is 1.58 mm. The average relative error of radius is 0.52%. The RMSE of radius is 1.59 mm. The average location deviation is 15.92 mm. The RMSE of location deviation is 2.66 mm.

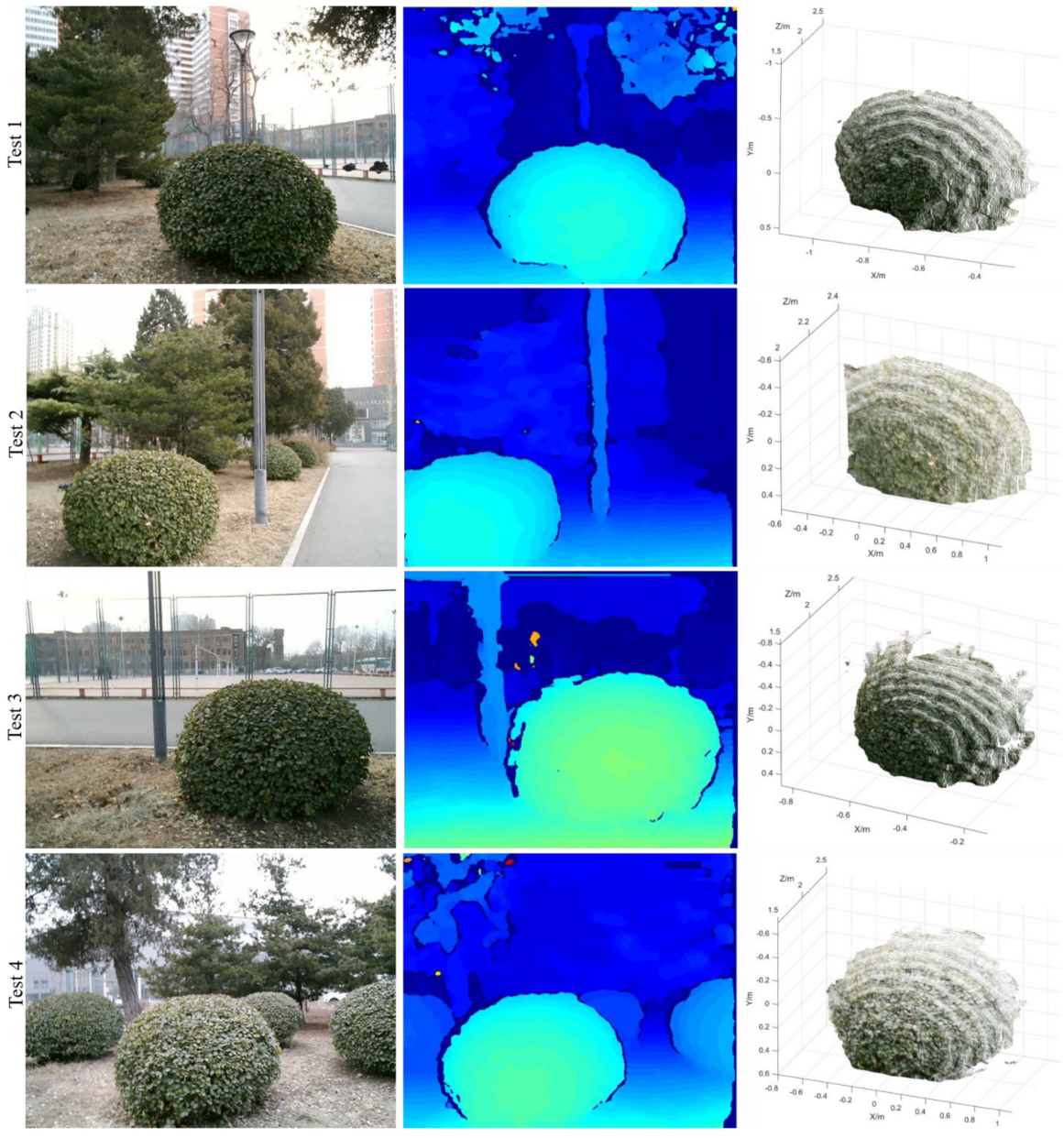


FIGURE 18 | 3D reconstruction result of outdoor test.

The pixel coordinates of p and q are marked as $p(x, y)$ and $q(k, l)$, respectively. Then, $G_{\sigma_s}(||\text{p-q}||)$ and $G_{\sigma_r}(I_p - I_q)$ can be expressed as

$$G_{\sigma_s}(||\text{p-q}||) = G_{\sigma_s}(x, y, k, l) = \exp\left(-\frac{(i-k)^2 + (j-l)^2}{2\sigma_s^2}\right) \quad (19)$$

$$G_{\sigma_r}(I_p - I_q) = G_{\sigma_r}(x, y, k, l) = \exp\left(-\frac{||I(i,j) - I(k,l)||^2}{2\sigma_r^2}\right) \quad (20)$$

where $I(i, j)$ and $I(k, l)$ are the disparity values of corresponding pixels in the disparity map.

Dimension Measurement of Spherical Hedges

According to the morphological characteristics of spherical hedges, the surface fitting model is established by the SGBM algorithm. The SGBM algorithm is a classic semi-global matching algorithm, which has the advantages of both stereo matching quality and processing rates.

After obtaining the disparity map through stereo matching, 3D point cloud coordinates of detected spherical hedges can be calculated by Equation 7. Then, the deformed shape of the spherical hedges is mapped and the error of coordinate and fitted coordinate of each 3D point is calculated. Finally, the coordinate of spherical hedges' center and its radius are obtained when the sum of error is minimal.

In the calculation process, $O(x_0, y_0, y_0)$ is the center of a fitting sphere, its corresponding radius is r, and (x_i, y_i, z_i) is the coordinate of a point cloud. The error formula of the actual coordinate and fitted coordinate of each 3D point can be expressed as (Guo et al., 2020)

$$e_i(x_0, y_0, z_0, r) = (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 - r^2 \quad (21)$$

Then, the sum of error is demonstrated as

$$E(x_0, y_0, z_0, r) = \sum_{i=1}^{N} e_i(x_0, y_0, z_0, r) \quad (22)$$

where N is the number of 3D point clouds, and E is the sum of errors.

In Equation 21, Eshows a function relation to x_0, y_0, z_0, and r. Thus, all the partial derivatives with respect to E are set to zero, and then, a minimum value of E can be obtained. The extreme value of partial derivative with respect to E can be expressed as

$$\frac{\partial E}{\partial x_0} = 0, \frac{\partial E}{\partial y_0} = 0, \frac{\partial E}{\partial z_0} = 0, \frac{\partial E}{\partial r} = 0 \quad (23)$$

With Equations 20–22 can be demonstrated as

$$\begin{cases} \sum_{i=1}^{N} e_i(x_i - x_0) = 0 \\ \sum_{i=1}^{N} e_i(y_i - y_0) = 0 \\ \sum_{i=1}^{N} e_i(z_i - z_0) = 0 \\ \sum_{i=1}^{N} e_i r = 0 \end{cases} \quad (24)$$

To solve out x_0, y_0, and z_0, Equation 23 can be transformed into

$$\begin{bmatrix} \overline{x^2} - \bar{x}^2 & \overline{xy} - \bar{x}\cdot\bar{y} & \overline{xz} - \bar{x}\cdot\bar{z} \\ \overline{xy} - \bar{x}\cdot\bar{y} & \overline{y^2} - \bar{y}^2 & \overline{yz} - \bar{y}\cdot\bar{z} \\ \overline{xz} - \bar{x}\cdot\bar{z} & \overline{yz} - \bar{y}\cdot\bar{z} & \overline{z^2} - \bar{z}^2 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} (\overline{x^3} - \bar{x}\cdot\overline{x^2}) + (\overline{xy^2} - \bar{x}\cdot\overline{y^2}) + (\overline{xz^2} - \bar{x}\cdot\overline{z^2}) \\ (\overline{x^2y} - \overline{x^2}\cdot\bar{y}) + (\overline{y^3} - \bar{y}\cdot\overline{y^2}) + (\overline{yz^2} - \bar{y}\cdot\overline{z^2}) \\ (\overline{x^2z} - \overline{x^2}\cdot\bar{z}) + (\overline{zy^2} - \bar{z}\cdot\overline{y^2}) + (\overline{z^3} - \bar{z}\cdot\overline{z^2}) \end{bmatrix} \quad (25)$$

where,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i, \overline{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i, \overline{xz} = \frac{1}{N}\sum_{i=1}^{N} x_i z_i,$$

$$\overline{yz} = \frac{1}{N}\sum_{i=1}^{N} y_i z_i, \overline{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2, \overline{y^2} = \frac{1}{N}\sum_{i=1}^{N} y_i^2, \overline{z^2} = \frac{1}{N}\sum_{i=1}^{N} z_i^2, \overline{x^2y} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 y_i,$$

$$\overline{x^2z} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 z_i, \overline{xy^2} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i^2, \overline{y^2z} = \frac{1}{N}\sum_{i=1}^{N} y_i^2 z_i, \overline{xz^2} = \frac{1}{N}\sum_{i=1}^{N} x_i z_i^2,$$

$$\overline{yz^2} = \frac{1}{N}\sum_{i=1}^{N} y_i z_i^2, \overline{x^3} = \frac{1}{N}\sum_{i=1}^{N} x_i^3, \overline{y^3} = \frac{1}{N}\sum_{i=1}^{N} y_i^3, \overline{z^3} = \frac{1}{N}\sum_{i=1}^{N} z_i^3.$$

Then, the radius of spherical hedges is obtained by

$$\overline{x^2} - 2x_0\bar{x} + x_0^2 + \overline{y^2} - 2y_0\bar{y} + y_0^2 + \overline{z^2} - 2z_0\bar{z} + z_0^2 = r^2 \quad (26)$$

RESULTS

Binocular Vision Calibration Test and Results

A calibration chessboard is applied in the experiment. The chessboard is placed in front of Camera-1 and Camera-2 with different positions and attitudes, and sixteen groups of images for calibration are captured. Then, the camera calibration toolbox (Toolbox_Calib) in MATLAB is used to extract corners in the chessboard. The detailed features of the chessboard are as follows: the material is armored glass; board size is 500 mm^2 × 500 mm^2; chessboard size is 390 mm^2 × 360 mm^2; check array is 13 × 12; check size is 30 × 30 mm^2; and the precision is ± 0.01 mm. In the captured calibration images, the number of corners that can

TABLE 3 | Outdoor test results of center coordinate and its radius.

No.	AR (mm)	MR (mm)	AER (mm)	RER (%)	AL (mm)	ML (mm)	LD (mm)
1	962.42	957.28	−5.14	0.53	(50, −80, 2,000)	(41.36, −67.50, 1,988.91)	18.84
2	881.92	877.89	−4.03	0.46	(150, −20, 2,150)	(152.35, −18.37, 2,165.26)	15.53
3	887.25	884.85	−2.4	0.27	(110, −150, 2,000)	(108.24, −145.38, 2,018.63)	19.27
4	932.75	937.26	4.51	0.48	(120, −80, 2,050)	(116.88, −73.21, 2,168.02)	19.51

AR, actual radius; MR, measured radius; AER, absolute error of radius; RER, relative error of radius; AL, actual location; ML, measured location; LD, location deviation. The average absolute error of radius is 4.02 mm. The average relative error of radius is 0.44%. The RMSE of radius is 1.01 mm. The average location deviation is 18.29 mm. The RMSE of location deviation is 1.61 mm.

be extracted from each image is 12 × 11. **Figure 13** shows one of the corner extraction results of Camera-2.

Taking the first corner in the lower left (marked in yellow square in **Figure 13** as the origin), the "X"-"Y" co-ordinate system is set up in a chessboard plane. The pixel coordinates of each corner can be obtained (Qiu and Huang, 2021). The world coordinates of corners are obtained based on the pixel coordinates of corners and check size. Then, the transformation matrix can be calculated by linear calculation. Additionally, by matrix decomposition, the intrinsic matrix (f_x, f_y, u_0 and v_0) of Camera-1 and Camera-2 can be obtained. In addition, a polynomial distortion correction model is built to correct the distortion, and the radial distortion coefficients and tangential distortion coefficients (k_1, k_2, p_1, and p_2) are given. The intrinsic parameters and distortion coefficients of Camera-1 and Camera-2 are listed in **Table 1**.

To test the calibration accuracy results listed in **Table 1**, the calibration errors of captured calibration images are analyzed, respectively. The coordinates of the corners in the "X"-"Y" co-ordinate system are obtained after back-projection and compared with the corresponding actual pixels of corners in the chessboard to obtain calibration errors. The binocular calibration errors of each image pair are shown in **Figure 14**. As can be seen in **Figure 14**, the binocular calibration errors for each pair of images are less than 0.05 pixels, and the average errors of Camera-1 and Camera-2 are both 0.04 pixels.

Then, the binocular vision calibration proceeds by using the binocular calibration toolbox in MATLAB. The installation of the two cameras is close to the coplanar and row alignment. As shown in **Figure 15**, the "1" and "2" represent the position and placing attitude of Camera-1 and Camera-2, respectively. The sixteen colored squares represent the positions and placing attitudes of the sixteen images of the calibration chessboard. In addition, the relative position between Camera-1 and Camera-2 can be obtained. Iterate over the intrinsic parameters and distortion coefficients of Camera-1 and Camera-2 obtained by monocular vision calibration. The transformation matrix and vector between Camera-1 and Camera-2 are given as $T = \begin{bmatrix} -119.2486 & 0.3206 & 3.3474 \end{bmatrix}^T$ and $R = \begin{bmatrix} 0.9998 & 0.0104 & -0.0111 \\ -0.0106 & 0.9998 & -0.0165 \\ 0.0109 & 0.0166 & 0.9998 \end{bmatrix}$.

The binocular calibration errors are also obtained by reverse projection of spatial coordinates of the corners, the binocular calibration errors for each pair of images are less than 0.07 pixels, and the average error of binocular calibration is less than 0.04 pixels. The calibration accuracy meets the requirements of the binocular vision system in this study.

Afterward, the images collected by this binocular vision system outdoor are used for stereo correcting, and the result is shown in **Figure 16**. The pixels of red dots from the top of the image are marked on the images. The pixels from the top of the original image of Camera-1 are 41, 267, and 428, whereas the values of Camera-2 are 39, 261, and 419, respectively. Herein, after stereo correction, the pixels of the same object in images of Camera-1 and Camera-2 are in the same row, and the pixels of makers after stereo correction are all 28, 264, and 428, respectively.

Laboratory Test and Results

To better reflect the 3D reconstruction effect of spherical hedges, a standard spherical hedge with a diameter of 60 mm was used to conduct a laboratory test first. The different test data sets could be obtained by changing the distances between the spherical hedges and the binocular vision system. Then, the stability and accuracy of this measurement system were verified according to the errors of the measured value and actual value. In the laboratory test, a straight line was marked in front of the binocular vision system, and seven different positions were set at the direction of Z by every 100 mm in range of 2,000–2,600 mm, described as red dots in **Figure 17**. Seven groups of images were captured, and the test values of the spherical center and its radius are shown in **Table 2**.

According to **Table 2**, the maximum and average error of radius of standard spherical hedges by the proposed system were 3.00 mm and 1.58 mm, respectively; maximum and average relative errors of radius were 1.00% and 0.52%, respectively; the root mean square error (RMSE) of the radius was 1.59 mm. Moreover, the relative error and error of radius increase with the distance in direction of Z, and the maximum relative error was 1.00% at the distance of 2,600 mm in direction of Z, which indicated the high monitoring accuracy and stability of the proposed system for radius measurement. The minimum, maximum, and average location deviations were 12.03, 19.12, and 15.92 mm in the range of 2,000–2,600 mm, and the RMSE of the center coordinate of spherical hedges was 2.66 mm. It showed that the proposed system had high accuracy in positioning and dimension measurement and had stability and applicability for different distances in a certain range.

Outdoor Test and Results

An outdoor test was conducted at China Agricultural University East Campus (Beijing, China). During the test, the weather was overcast and the leaves of spherical hedges were slightly yellow and sparse. A number of four spherical hedges were randomly selected on the campus; therefore, the results have a certain generality. The spherical hedges were non-standard spheres and their radius was unknown; therefore, for each spherical hedge, six groups of images were captured at different positions. The distances between the proposed system and spherical hedges were all around 2,000 mm. The outdoor scene image acquired by the left camera, the disparity map obtained by stereo matching, and the 3D shape reconstruction image of the proposed system are shown in **Figure 18**.

In the outdoor test, the actual center position and radius of spherical hedges were measured manually using a tap. In each test, the actual radius was collected manually by six different positions, and the average value was determined. The results of the center coordinate and its radius in **Figure 18** are shown in **Table 3**.

According to **Table 3**, the maximum and average errors of the radius of measured spherical hedges in the outdoor test were 5.14 and 4.02 mm, respectively; maximum and average relative errors of radius were 0.53% and 0.44%; the and RMSE of the radius was 1.01 mm, respectively. At the distance of around 2,000 mm in direction of Z, the maximum and average location deviation were 19.51 and 18.29 mm, respectively. It indicated a high measurement accuracy and stability of the proposed system for outdoor sphere center positioning and radius detection.

DISCUSSION

A binocular vision system for spherical hedge reconstruction and measurement was proposed in this work to provide front-end visual information for pruning robots. Through theoretical analysis and experimental verification, this shape reconstruction and dimension measurement method showed high accuracy in both spherical center positioning and radius measurement. The conclusions of this study were as follows:

(1) The binocular vision platform was built based on the theory of binocular parallel structure. After binocular camera calibration, stereo image correcting was used based on Bouguet's algorithm to improve the accuracy of shape reconstruction. Meanwhile, the captured 2D images were processed through filtering algorithm, segmentation, edge extraction, etc. Then, an improved SGBM algorithm was applied to obtain a good disparity map.

(2) The sharp reconstruction and measurement method were tested in a laboratory and outdoors in the detection range of 2,000–2,600 mm. The laboratory test result showed that the average error and average relative error of standard spherical hedges radius were 1.58 mm and 0.53%, respectively; the average location deviation of the center coordinate of spherical hedges was 15.92 mm in range of 2,000–2,600 mm. The outdoor test showed that the average error and average relative error of spherical hedges radius by the proposed system were 4.02 mm and 0.44%, respectively; the average location deviation of the center coordinate of spherical hedges was 18.29 mm. Therefore, the proposed system could be employed for the visual information acquisition of various trimming robots due to its excellent applicability.

Future studies may involve expanded tests on different shapes of hedges to clarify the accuracy and stability of the proposed system further. This study provides key technical support for visual detection in studies of trimming robots.

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/s.

AUTHOR CONTRIBUTIONS

YZ, JG, TR, and HL built the system, conducted the experiments, and wrote the manuscript. BZ, JZ, and YY designed the measurement method. All authors discussed the measurement method and designed the laboratory and outdoor experiments.

FUNDING

This work was supported by the National Key Research and Development Project of China (2019YFB1312305) and the Modern Agricultural Equipments and Technology Demonstration and Promotion project of Jiangsu Province (NJ2020-08). We would like to thank a team of the Beijing Research Center of Intelligent Equipment for Agriculture for financial support of the research.

ACKNOWLEDGMENTS

We would like to thank Taiyu Wu, Lin Gui, Xiao Su, and Jun Liu for helpful discussion and assistance of the research.

REFERENCES

Caraffa, L., Tarel, J. P., and Charbonnier, P. (2015). The guided bilateral filter: when the joint/cross bilateral filter becomes robust. *IEEE Trans. Image Proc.* 24, 1199–1208. doi: 10.1109/TIP.2015.2389617

Guo, X., Shi, Z., Yu, B., Zhao, B., Li, K., and Sun, Y. (2020). 3D measurement of gears based on a line structured light sensor. *Precision Eng.* 61, 160–169. doi: 10.1016/j.precisioneng.2019.10.013

Hong, P. N., and Ahn, C. W. (2020). Robust matching cost function based on evolutionary approach. *Exp. Syst. Appl.* 161:113712.

Ji, W., Meng, X., Qian, Z., Xu, B., and Zhao, D. (2017). Branch localization method based on the skeleton feature extraction and stereo matching for apple harvesting robot. *Int. J. Adv. Robotic Syst.* 14:276. doi: 10.1177/1729881417705276

Jin, Z., Sun, W., Zhang, J., Shen, C., Zhang, H., and Han, S. (2020). Intelligent tomato picking robot system based on multimodal depth feature analysis method. *IOP Conf. Ser. Earth Environ. Sci.* 440:74. doi: 10.1088/1755-1315/440/4/042074

Kaljaca, D., Mayer, N., Vroegindeweij, B., Mencarelli, A., Henten, E. V., and Brox, T. (2019a). "Automated boxwood topiary trimming with a robotic arm and integrated stereo vision," in *Proceeding of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Macau. (China)*.

Kaljaca, D., Vroegindeweij, B., and Henten, E. (2019b). Coverage trajectory planning for a bush trimming robot arm. *J. Field Robot.* 37, 283–308. doi: 10.1002/rob.21917

Li, Z., Xu, E., Zhang, J., Meng, Y., Wei, J., Dong, Z., et al. (2022). AdaHC: adaptive hedge horizontal cross-section center detection algorithm. *Comput. Electr. Agric.* 192:106582. doi: 10.1016/j.compag.2021.106582

Lili, W., Bo, Z., Jinwei, F., Xiaoan, H., Shu, W., Yashuo, L., et al. (2017). Development of a tomato harvesting robot used in greenhouse. *Int. J. Agric. Biol. Eng.* 10, 140–149. doi: 10.25165/j.ijabe.20171004.3204

Long, L., and Dongri, S. (2019). "Review of camera calibration algorithms," in *Advances in Computer Communication and Computational Sciences*, eds S. Bhatia, S. Tiwari, K. Mishra, and M. Trivedi (Singapore: Springer), 723–732. doi: 10.1007/978-981-13-6861-5_61

Qiu, Z.-C., and Huang, Z.-Q. (2021). A shape reconstruction and visualization method for a flexible hinged plate using binocular vision. *Mech. Syst. Signal Proc.* 158:107754. doi: 10.1016/j.ymssp.2021.107754

Romaniuk, R. S., and Roszkowski, M. (2014). "Optimization of semi-global stereo matching for hardware module implementation," in Proceedings of the Symposium on Photonics Applications in Astronomy, Communications, Industry and High-Energy Physics Experiments, Warsaw. doi: 10.1117/12.2075012

Strisciuglio, N., Tylecekg, R., Petkova, N., Biberb, P., Hemmingc, J., Hentenc, E., et al. (2018). *TrimBot2020: An Outdoor Robot for Automatic Gardening*. Munich: Vde Verlag publisher.

Wu, W., Zhu, H., and Zhang, Q. (2017). Epipolar rectification by singular value decomposition of essential matrix. *Multi. Tools Appl.* 77, 15747–15771. doi: 10.1007/s11042-017-5149-0

Yang, Q., Chang, C., Bao, G., Fan, J., and Xun, Y. (2018). Recognition and localization system of the robot for harvesting hangzhou white chrysanthemums. *Int. J. Agric. Biol. Eng.* 11, 88–95. doi: 10.25165/j.ijabe.20181101.3683

Ye, L., Duan, J., Yang, Z., Zou, X., Chen, M., and Zhang, S. (2021). Collision-free motion planning for the litchi-picking robot. *Comput. Electr. Agric.* 185:106151. doi: 10.1016/j.compag.2021.106151

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fpls.2022.881560

Design of a Phenotypic Sensor About Protein and Moisture in Wheat Grain

*Yiming Liu[1,2,3,4], Donghang Li[1,2,3,4], Huaiming Li[1,2,3,4], Xiaoping Jiang[1,2,3,4], Yan Zhu[1,2,3,4], Weixing Cao[1,2,3,4] and Jun Ni[1,2,3,4]**

[1]College of Agriculture, Nanjing Agricultural University, Nanjing, China, [2]National Engineering and Technology Center for Information Agriculture (NETCIA), Nanjing, China, [3]Engineering Research Center of Smart Agriculture, Ministry of Education, Nanjing, China, [4]Collaborative Innovation Center for Modern Crop Production Co-sponsored by Province and Ministry, Nanjing, China

A near-infrared (NIR) spectrometer can perceive the change in characteristics of the grain reflectance spectrum quickly and nondestructively, which can be used to determine grain quality information. The full-band spectral information of samples of multiple physical states can be measured using existing instruments, yet it is difficult for the full-band instrument to be widely used in grain quality detection due to its high price, large size, non-portability, and inability to directly output the grain quality information. Because of the above problems, a phenotypic sensor about grain quality was developed for wheat, and four wavelengths were chosen. The interference of noise signals such as ambient light was eliminated by the phenotypic sensor using the modulated light signal and closed sample pool, the shape and size of the incident light spot of the light source were determined according to the requirement for collecting the reflectance spectrum of the grain, and the luminous units of the light source with stable light intensity and balanced luminescence were developed. Moreover, the sensor extracted the reflectance spectrum information using a weak optical signal conditioning circuit, which improved the resolution of the reflectance signal. A grain quality prediction model was created based on the actual moisture and protein content of grain obtained through Physico-chemical analyses. The calibration test showed that the R^2 of the relative diffuse reflectance (RDR) of all four wavelengths of the phenotypic sensor and the reflectance of the diffusion fabrics were higher than 0.99. In the noise level and repeatability tests, the standard deviations of the RDR of two types of wheat measured by the sensor were much lower than 1.0%, indicating that the sensor could accurately collect the RDR of wheat. In the calibration test, the root mean square errors (RMSE) of protein and moisture content of wheat in the Test set were 0.4866 and 0.2161%, the mean absolute errors (MAEs) were 0.6515 and 0.3078%, respectively. The results showed that the NIR phenotypic sensor about grain quality developed in this study could be used to collect the diffuse reflectance of grains and the moisture and protein content in real-time.

Keywords: sensor, quality phenotype, near-infrared, multi-source circular structure, Fresnel reflectance concentrator, optical simulation, feedback driver, neural network modeling

OPEN ACCESS

Edited by:
Daobilige Su,
China Agricultural University, China

Reviewed by:
Aichen Wang,
Jiangsu University, China
Abbas Atefi,
California Polytechnic State University, United States

****Correspondence:***
Jun Ni
nijun@njau.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *22 February 2022*
Accepted: *01 April 2022*
Published: *06 May 2022*

Citation:
Liu Y, Li D, Li H, Jiang X, Zhu Y, Cao W and Ni J (2022) Design of a Phenotypic Sensor About Protein and Moisture in Wheat Grain. Front. Plant Sci. 13:881560. doi: 10.3389/fpls.2022.881560

INTRODUCTION

Real-time non-destructive testing of grain quality can lower the cost of field management by providing reference information for fertilization in the grain production process. It can also provide information for quality-based pricing during trade and provide data for grain classification to improve the efficiency of grain processing. However, traditional grain quality detection methods mainly rely on indoor Physico-chemical analysis. For instance, the grain moisture content was determined by the direct drying method (The National Health and Family Planning Commission of the People's Republic of China, 2016), and the protein content was determined by the Kjeldahl method (The National Health and Family Planning Commission of the People's Republic of China and State Food and Drug Administration, 2010). These methods cannot be applied to real-time quality detection in the process of grain circulation due to destructive sample preparation, complex analysis process, and poor timeliness (El-Mesery et al., 2019; Hussain et al., 2019).

Near-infrared (NIR) spectroscopy has developed rapidly as an efficient, green, and nondestructive analysis and detection method in recent years. The response characteristics of the NIR reflectance spectrum of grain are closely related to grain quality. The quality of grains can be described quantitatively using the NIR wavelength that is sensitive to the quality content (Wang, 2010; Santos et al., 2013; Zhu et al., 2015; Caporaso et al., 2018). By collecting the NIR transmitted spectrum of six types of single-grain wheat, a prediction model for NIR transmitted spectrum and the protein content of corresponding single-grain wheat was proposed by Delwiche (1995) based on partial least square. Using the Foss InfraXact™ Lab/Pro spectrometer (570–1850 nm), Arazuri et al. (2012) collected the spectrum of wheat in the NIR area, analyzed the relationships of rheological parameters between spectrum and tenacity, extensibility, deformation energy of wheat, which provided technical support for obtaining the rheological parameters of wheat during harvest and transportation of wheat. In a study by Li et al. (2013). a spectrometer was used to collect the transmitted spectrum (840–1,048 nm) of brown rice. This information was used to accurately determine the moisture and protein content of brown rice in the grain elevator. In summary, the above studies on NIR of grain showed that the quality information, e.g., moisture, starch, and protein content of grains (e.g., wheat and rice) can be obtained quantitatively.

Based on the NIR spectroscopy technology, grain quality detection devices have been developed. Zhang et al. designed a NIR analysis system with a wavelength of 800–1,100 nm based on a charge-coupled device and fixed optical grating, and a wheat moisture and protein model was created using the PLS method. However, this system was complex and large, it could not be used for real-time on-site detection (Zhang et al., 2009). Hidaka et al. developed a NIR reflectance spectrometer of 740–1,140 nm using the halogen lamp and grating dispersion, which was integrated with the harvester to determine the protein content of brown rice. However, when tested in the field, the correlation coefficient R of the detection value and the real value of protein was only 0.65 since the system did not take into account the influence of complex environment factors (Hidaka et al., 2011). In a study by Wen et al. light-emitting diodes (LEDs) and 14 narrow-band interference filters with a wavelength between 800 and 1,100 nm were used to develop a single-grain wheat composition analyzer, which realized the real-time detection of wheat protein content. However, the luminous efficiency of the LED of 950 nm was very low, at 1,020–1,050 nm, and the measurement error under repeated sample loading conditions was largely due to the light spot difference caused by the position of the light source (Wen and Ji, 2004). Wu et al. used a 6-row-8-column full-enclosed LED structure, combined with a FLAME-NIR spectrometer (900–1,700 nm) manufactured by Ocean Insight (Orlando, FL, United States), to design an instrument for detecting the protein content of single-grain wheat. This instrument requires a large number of light sources and has large power consumption, and it can only be used after the light source has been preheated for 30 min, thus the timeliness of the instrument was not ideal (Wu et al., 2018).

Studies on the analytical techniques based on NIR spectroscopy have played an important role in grain quality detection. However, current studies are still facing the following problems.

1. Commercial instruments mostly consist of spectrometers operating in the full spectral band. They are complex, bulky, expensive to produce, difficult to adapt to different detection requirements, and do not have a grain quality detection model; thus, they are incapable of outputting grain quality information in real-time.
2. Most instruments made in-house consist of spectrometers operating in characteristic spectral bands. The incident light spots of multiple wavelength light sources suffer from the problem of inconsistent detection regions. For grains with uneven distributions of nutritional quality, inconsistent detection regions will result in systematic errors, thereby rendering the instrument unsuitable for grain quality detection.

In this study, a NIR phenotypic sensor about protein and moisture in wheat grain (PSPMWG) was developed. Compared with previous studies, this study made contributions in the following aspects.

1. A NIR PSAGQ was developed, which had a simple structure, was lightweight and was able to detect the grain quality rapidly, nondestructively in real-time. The sensor collects the diffuse reflectance spectrum in real-time and couples the quality detection model to generate grain quality data immediately. In addition, this study expounds on the optical system design and control system development process of the phenotypic sensor in detail, which lays a good foundation for the secondary development of the phenotypic sensor and the application of online detection.
2. A circular structure of multiple light sources was proposed, where the incident light spots of the multiple light sources were consistent, which reduced the measurement error under repeated sample loading conditions caused by the differences in incident light spots. Different incident light spots could

Abbreviations: PSPMWG, phenotypic sensor about protein and moisture in wheat grain; RDR, relative diffuse reflectance.

be formed by multiple light sources on the detection plane. The overlapping circular part was intercepted to ensure the consistency of incident light spots from multiple light sources and to improve the spectral accuracy of the sensor.

3. A detection model of wheat protein and moisture was built, which was able to output the grain quality information in real-time after obtaining the grain spectrum. The model could be directly used in multiple instruments through the self-correcting luminous state by correcting the reflectance of the sensor through the diffusion fabric and correcting the sample pool. The instrument could be used during grain production, trading, and grain processing.

MEASUREMENT PRINCIPLES OF THE SENSOR

There are two methods for measuring the NIR: the transmission method (Delwiche et al., 1996; Li et al., 2013) and the reflection method (Hidaka et al., 2011; Priya et al., 2015). The transmission method applies to uniform or transparent samples, and the NIR absorbs the same proportion of light on each equivalent thickness medium on the optical path, that is, the absorbed amount of light is in direct proportion to the number of molecules that generate light absorption on the optical path. The reflection method is mostly used for solid samples, and the NIR passes through a tortuous and irregular optical path in the solid body. The absorbance is related to the scattering coefficient and absorption coefficient of the sample.

Wheat grain is the research object of this study. Its nutritional components are unevenly distributed in the grain. The PSPMWG can obtain the nutritional quality information of the wheat through the grain spectrum measured using the diffuse reflection method.

The relative diffuse reflectance (RDR) is defined as:

$$R = I / I_0 \quad (1)$$

where I is the diffuse reflection light intensity of the measured sample, and I_0 is the diffuse reflection light intensity of the background.

Diffuse reflectance spectral signals can also be characterized by absorbance, as defined by

$$A = \log \frac{1}{R} \quad (2)$$

This work used RDR to characterize the diffuse reflectance spectrum. When measuring the diffuse reflectance spectrum, each measurement result of the sample is affected by the loading conditions. Because the internal arrangement and distribution of the samples loaded each time are different, the optical path will change, which will lead to spectral changes and poor spectral repeatability. With the design of a closed sample cell and annular light sources at multiple spectral bands, a NIR PSPMWG could reduce the influence of sample loading conditions on spectral measurement and improve spectral repeatability.

Due to the low manufacturing cost and strong environmental adaptability of spectrometers working in the shortwave infrared and NIR regions (780–1,100 nm; Risius et al., 2014), we used a NIR light source with wavelengths sensitive to grain quality in the shortwave infrared and NIR regions. **Figure 1** shows the NIR spectra (570–1,100 nm) of 300 winter wheat samples collected by a Foss InfraXact™ Lab/Pro spectrometer (Bec et al., 2020). The regions where the NIR spectral characteristics of the wheat varied were marked by red boxes in the figure. In addition, according to previous studies, (Zhang et al., 2009; Hidaka et al., 2011; Arazuri et al., 2012; Li et al., 2013) on the wavelengths sensitive to grain moisture and protein content, four NIR LEDs with wavelengths of 780, 910, 980, and 1,050 nm were used as the light sources for portable quality detection.

DESIGN OF THE SENSOR

Overall Design

The NIR PSPMWG is composed of an optical system (including a light-emitting unit and a spectral collection unit) and a control system (including hardware circuits and software). The light-emitting unit generates incident spectra, the spectral collection unit obtains the RDR of the grain, and the hardware circuit is used to drive the light-emitting unit and the spectral collection unit. The software couples the RDR with the grain quality detection model to determine the quality content values, which are then displayed on the liquid-crystal display (LCD) with an audible output. The overall structure of the sensor is shown in **Figure 2**.

Optical System

Overall Design of the Optical System

A stable and reliable optical system is very important for the NIR PSPMWG to obtain the diffuse reflectance spectrum accurately. The light-emitting unit provides a stable and reliable incident spectrum through NIR light sources, and the spectral collection unit collects the diffuse reflectance spectrum of the grain accurately *via* the photoelectric detector.

Design of the Light-Emitting Unit

To obtain stable reflectance spectrum signals during grain quality detection, the sensor needs to have light sources with large luminous intensity and stability. Common NIR light sources include a halogen lamp and LED. The halogen lamp is bright and has a long service life, yet it is large and has high power consumption and poor stability of luminous intensity. In addition, the halogen lamp has poor timeliness because it has to be preheated for 30 min before use (Liu et al., 2019). The LED has a long service life, low power consumption, compact dimension, and affordability (Bec et al., 2020). In this study, after comparing the advantages and disadvantages of the two types of light sources, the LED was chosen as the light source for the NIR PSPMWG, with a half-value angle of 12°.

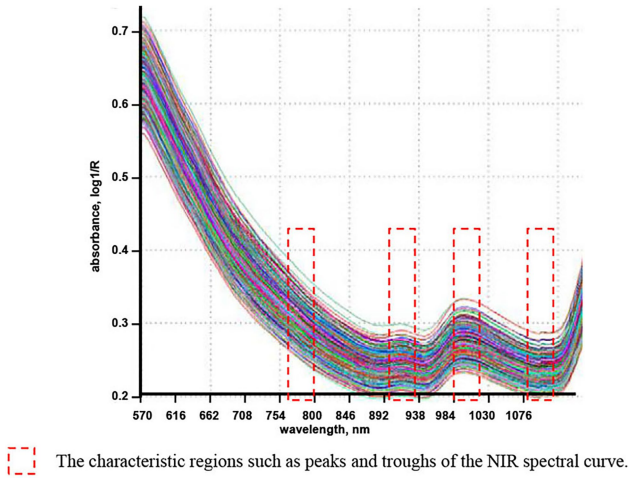


FIGURE 1 | The NIR spectrum of the wheat.

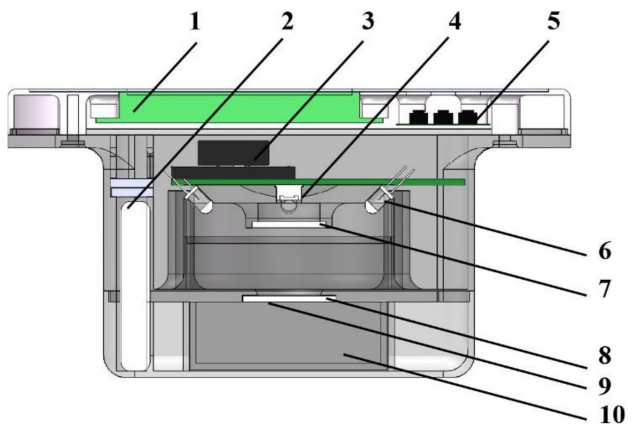


FIGURE 2 | The overall design of the NIR PSPMWG. 1. liquid crystal display (LCD) 2. battery 3. voice broadcast module 4. photoelectric detector 5. button 6. NIR LED 7. Fresnel lens 8. beam shaping diffusion film 9. detection window plane 10. sample pool.

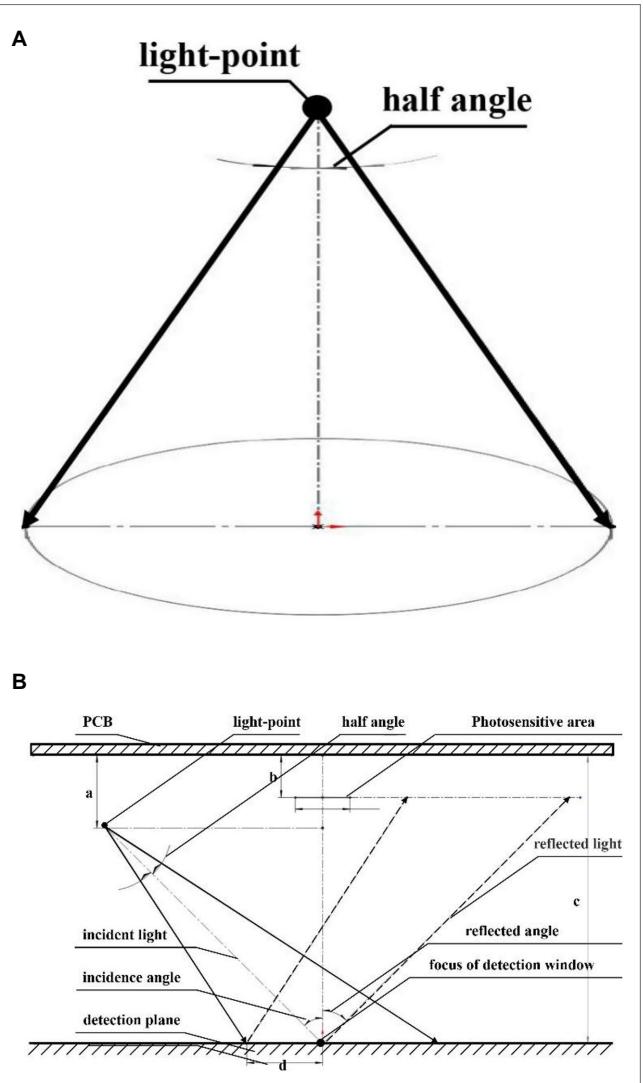


FIGURE 3 | **(A)** Emission light spot of a single LED, **(B)** Emission light spot of LED in the light-emitting unit.

The light-emitting unit was designed based on the selected NIR LEDs, the light-emitting path of the light source is shown in **Figure 3A**. A circular spot was formed on the vertical plane. The farther away from the light-emitting point, the weaker the light intensity of the spot. Therefore, under a given power consumption, the efficiency of the emission spot could be maximized when the light-emitting point of the light source, the light-sensitive surface center point of the photoelectric detector, and the center point of the detection window plane was on the same plane. In this way, the photoelectric detector could accurately detect the diffuse reflectance spectrum. Moreover, to reduce the influence of the specular reflection light on the collection of reflectance spectrum and to ensure that the specular reflection light is not collected by the photoelectric detector, the incident angle of the LED light was 45°. The design of the light-emitting unit is shown in **Figure 3B**.

According to the above design, the incident light spots of the four light sources are formed in different areas of the detection plane, leading to system errors. Thus, a circumferential layout with even distribution of four LEDs was used (**Figure 4A**). The incident light spots generated by the four light sources on the detection plane are shown in **Figure 4B**. The incident light spot areas formed by the four LEDs were different, yet there were some overlapping areas. The overlapping areas of the four incident light spots were chosen as the detection window to ensure consistency in the detection area of the four light sources and to reduce system error. The nutritional components of wheat were distributed unevenly, thus there might be random errors in the detection of single-grain wheat. There should be multiple wheat grains in the detection area so that the diffuse reflectance spectrum can effectively lower random errors. The dimensions of the detection window are shown in **Figure 4C**. When the detection window is a circle

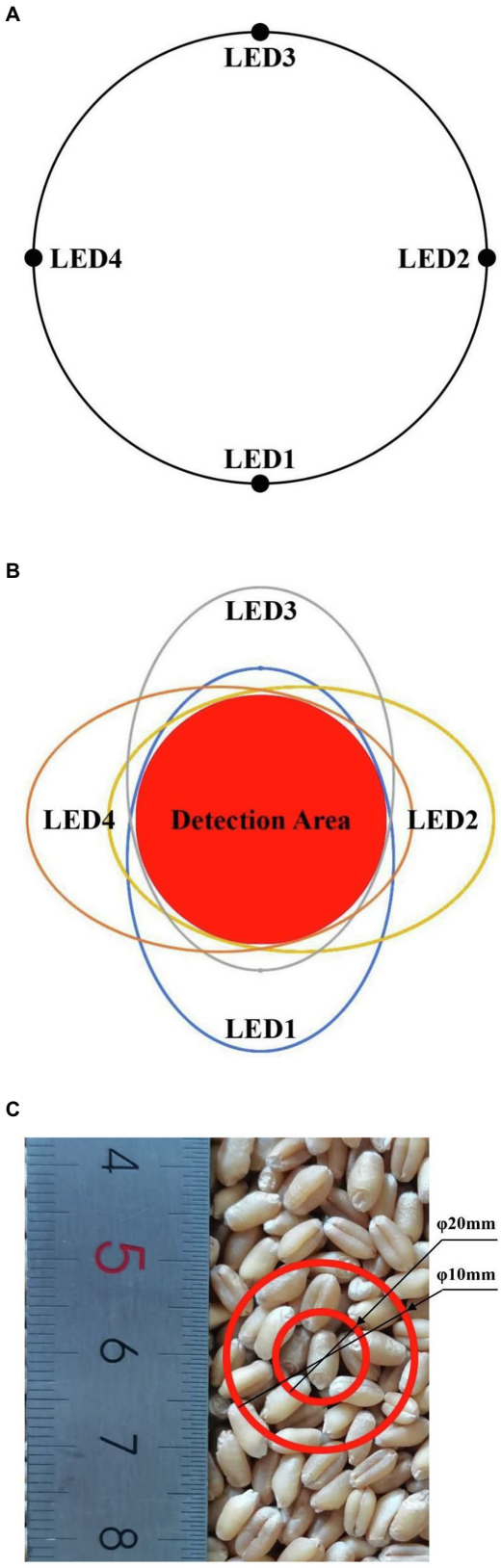


FIGURE 4 | (A) The circular layout of four LED light sources, **(B)** Incident light spots and detection window plane, **(C)** Size of the detection window.

with a diameter of 10 mm and there was only one single grain of wheat, there might be a random error. When the detection window was a circle with a diameter of 20 mm and there were 10 grains in the window, the random error seen in single-grain wheat detection was lowered effectively. Therefore, the detection window was set to a circle with a diameter of 20 mm, and the optical path and grain were separated by quartz glass. The specific design parameters of the light-emitting unit are shown in **Figure 5**.

The light-emitting unit was simulated and analyzed by the TracePro optical simulation software to verify the consistency of the incident light spots by the four light sources. First, four light sources with an incident angle of 45° and a half-value angle of 12° were constructed, and a light beam blocking plate was placed at the appropriate location to substitute the detection plane in the actual design. The simulation results of the light-emitting system are shown in **Figure 6A**, where the bottom view showed that the elliptical light spots of the four light sources formed a circular light coincidence area on the detection surface (the same sample area detected with the four light sources). This circular area was used as the detection window to ensure the consistency of the incident light spots from the four light sources. The irradiance analysis diagram

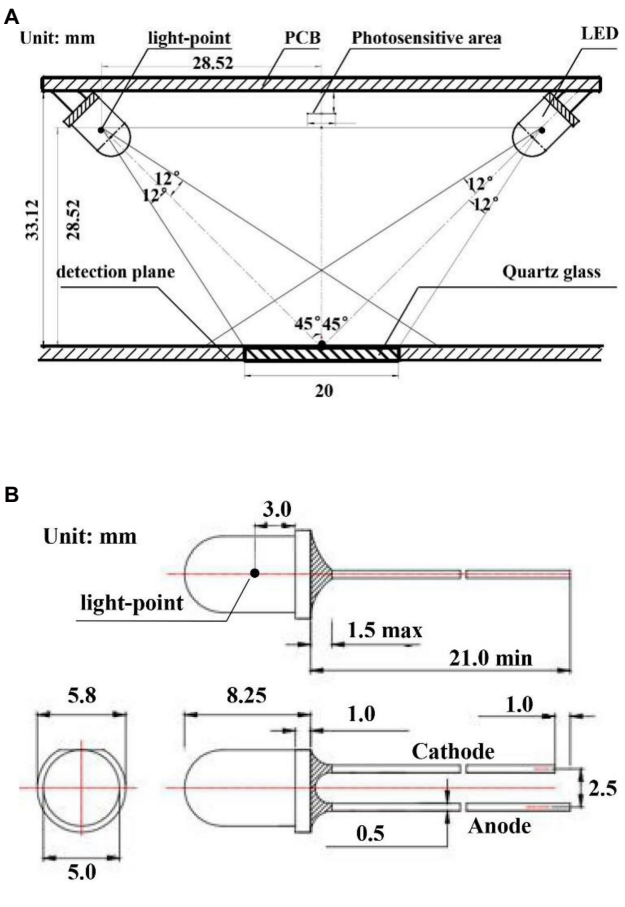


FIGURE 5 | (A) Design parameters of the light-emitting unit, **(B)** Outer dimension of LEDs.

from the light beam blocking plate (detection plane) is shown in **Figure 6B**, where the location of the detection window is marked by a red circle in the figure.

Design of the Spectral Collection Unit

A photoelectric detector is an electronic component that converts optical signals to electric signals (Yang et al., 2005), which ensures the accurate collection of the diffuse reflectance spectrum. The selection of a NIR photoelectric detector is mainly determined by three factors: response scope, response speed, and sensitivity. In this study, the spectral response of the silicon photoelectric detector ranged from 320 to 1,100 nm, the response time was 3.6 μs, and the sensitivity for the four wavelengths was above 0.4 A/W. These parameters satisfied the requirement of the NIR PSPMWG. The light emitted from the four LEDs could be converted into current by the photodiode quickly and accurately.

The spectral collection unit was designed based on the selected photoelectric detector. The diffuse reflectance spectrum emitted from the detection window was a circle with a diameter of 20 mm, while the photosensitive area of the selected photoelectric detector was a $3.6 \times 3.6\,mm^2$. The collection of diffuse reflectance spectrum by the photoelectric detector is shown in **Figure 7**. The photoelectric detector only collects the diffuse reflectance spectrum that is received by the photosensitive area; it does not collect the other diffuse reflectance spectrum, leading to low accuracy in the diffuse reflectance spectrum.

To improve the collection accuracy in the diffuse reflectance spectrum, all diffuse reflectance spectrum of grain has to be received by the photoelectric detector, that is, the circular diffuse reflectance spectrum with a diameter of 20 mm needs to be reflected on the square photosensitive area. Therefore, it is necessary to add a condenser between the detection window and the photoelectric detector. The Fresnel lens is developed based on the Fresnel theory (Wu et al., 2018), it is small and light and is suitable for the application of the sensor. The diffuse reflectance spectrum is condensed as much as possible without blocking the incident light, and the Fresnel lens has the same dimensions as the detection window, i.e., a circle with a diameter of 20 mm. The schematic diagram of the photosensitive optical path is shown in **Figure 8**.

The calculation of the focal length of the lens is as follows:

$$\frac{f}{d} = \frac{H}{D} \tag{3}$$

where D is the dimension of the reflection spot, d is the dimension of the photosensitive element, and H is the distance from the light spot to the lens.

Based on the dimensions of the spectrum collection unit, the focal length of the Fresnel lens was 15 mm, and the diameter was 20 mm, with a thickness of 2 mm. The above parameters were substituted to the calculation formula (Yao et al., 2020) of Fresnel lens to calculate the intersection angle between the lens sawtooth and the vertical direction, which was 2.67°. The spectral collection unit is shown in **Figure 9**.

The spectrum acquisition unit was simulated and analyzed by the TracePro optical simulation software to verify the light-condensing performance of the Fresnel lens used in the system.

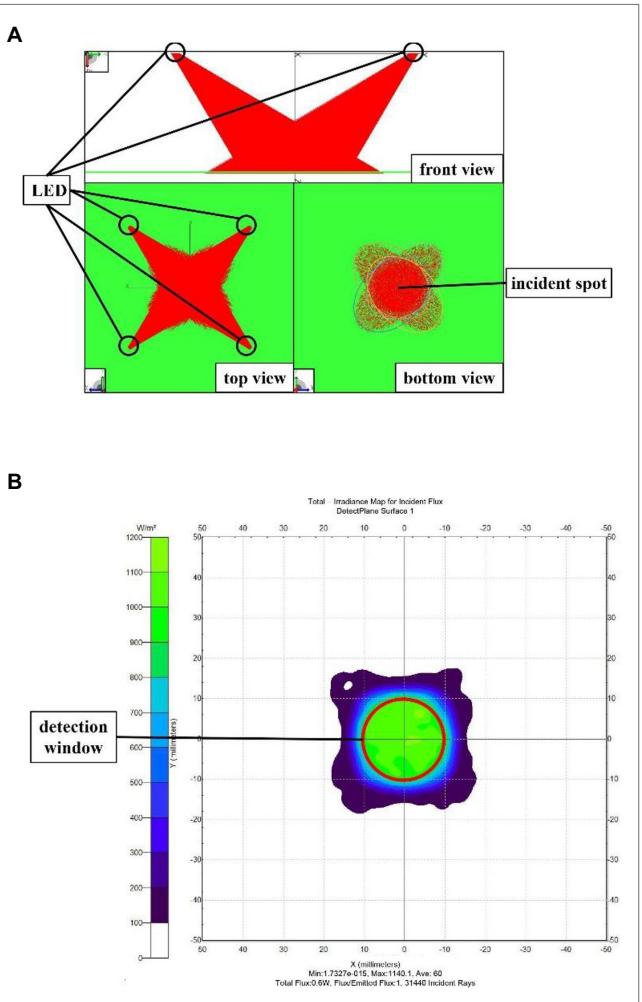


FIGURE 6 | (A) TracePro simulation of the light-emitting unit, **(B)** Irradiance analysis of the detection plane.

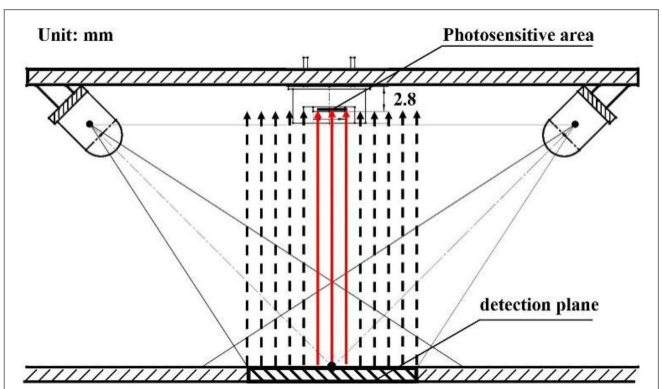


FIGURE 7 | Direct spectrum acquisition of diffuse reflection tube.

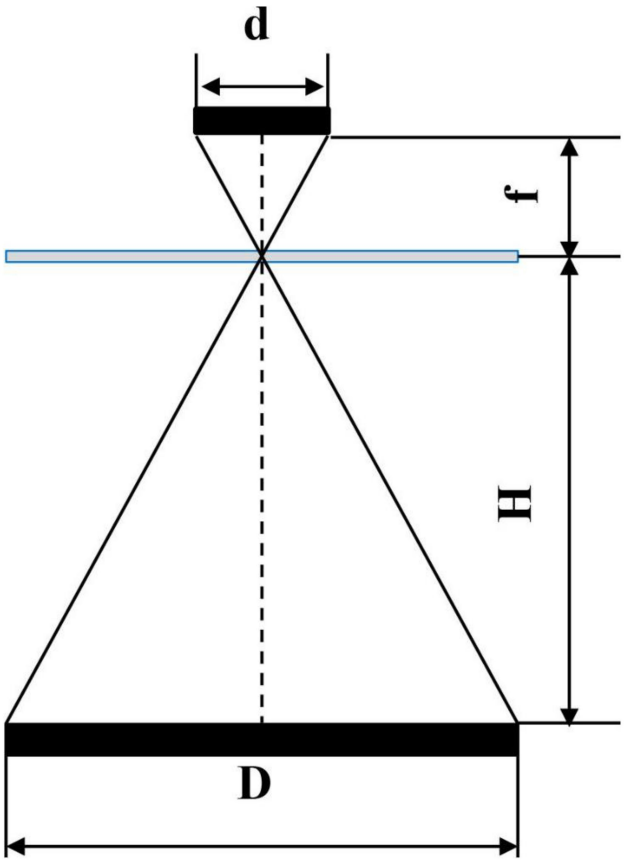


FIGURE 8 | Inductive light.

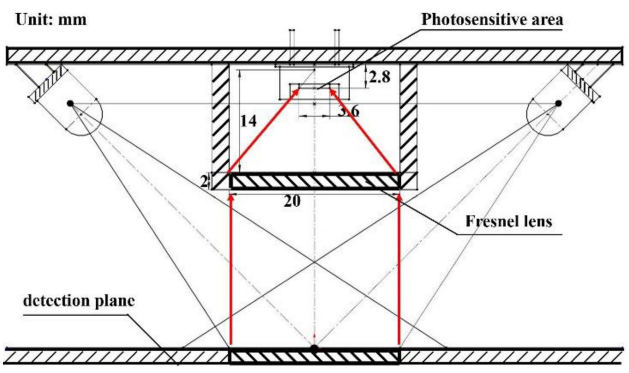


FIGURE 9 | Spectral acquisition unit design.

The Fresnel lens was constructed based on a theoretical design, and a light-blocking plate was placed at an appropriate location to replace the photosensitive plane in the actual design. Then, the light-condensing performance of the lens was assessed by a parallel light beam source. The simulation results of the spectrum acquisition unit are shown in **Figure 10A**. The parallel light beam was converged by the lens onto the light-blocking plate (photosensitive plane), with most of the light falling

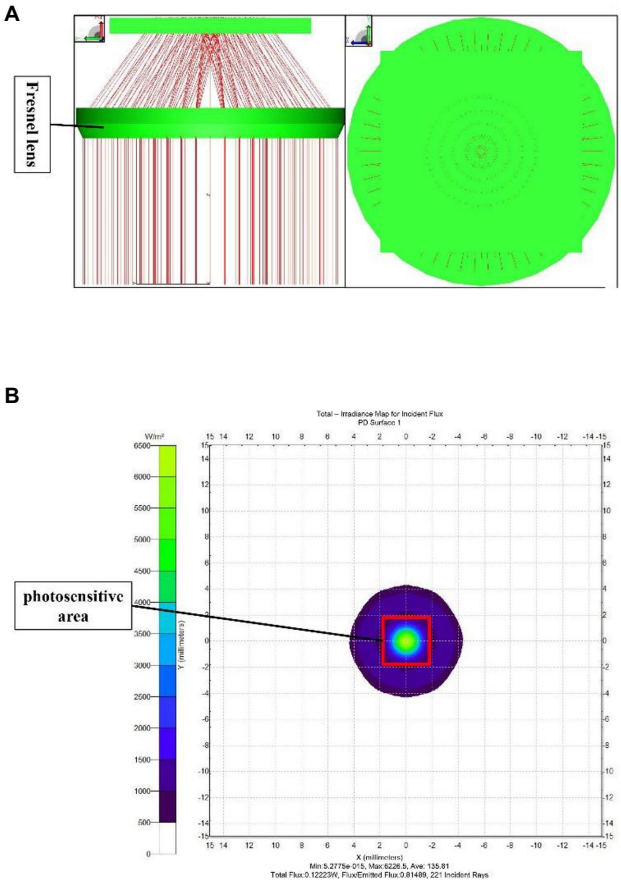


FIGURE 10 | (A) TracePro simulation of the spectrum acquisition unit, **(B)** Irradiance analysis diagram of the photosensitive plane.

within an area of 3.6 × 3.6 mm2. The photodetector placed in this area could accurately collect the convergent diffuse reflection spectrum. The irradiance analysis diagram of the light-blocking plate (photosensitive plane) is shown in **Figure 10B**, where the position of the photosensitive plane of the photosensitive element is marked by a red box in the figure.

Control System

Overall Design of the Control System

The control system consisted of hardware circuitry and a software system. The hardware circuit is composed of the light source driver circuit, the spectral collection circuit, and the signal conditioning circuit. The hardware circuit is mainly used to ensure stable luminescence of the light source and accurate collection of the diffuse reflectance spectrum of grain. The software system mainly achieves the collection of the diffuse reflectance spectrum of grain, couples that data with the grain quality prediction model, and outputs the grain quality results.

Hardware Circuit Design

The luminous intensity of the light-emitting unit needs to be stable for the PSPMWG to accurately collect the diffuse

reflectance spectrum. The luminous intensity of the light-emitting unit is affected by its factors (e.g., light source loss and heat dissipation) and driver factors such as drive current. The light source is driven by the optical power feedback circuit. The light source driver circuit based on optical power feedback mainly consists of four parts: optical power feedback, comparative amplification, constant current drive, and compensation filter. The circuit diagram is shown in **Figure 11**.

When the luminous power of the light source under constant current reduces, the feedback voltage in the optical power feedback circuit decreases, the drive voltage in the comparative amplification part increases, the drive current in the constant current driving part increases, and the luminous power of the driving light source increases, and vice versa. In the compensation filter part, the light source drive signal is stabilized by reducing the ringing and overshoot driven by the modulation signal through the resistive–capacitive compensation filter circuit. The drive current and luminous power of the light source can be balanced through the optical power feedback to achieve stable luminous power of the light source.

The accuracy of the grain diffuse reflectance spectrum obtained by the spectrum collection circuit directly affects the accuracy of grain quality data. In the spectrum collection circuit, the diffuse reflectance spectrum is first converted into light current through the photoelectric detector, which is converted into a voltage signal using the current–voltage conversion circuit, then the noise signal is filtered. Lastly, the voltage signal is amplified to the collection range of the microprocessor for further processing and output.

The spectrum collection circuit mainly consists of three parts: current–voltage conversion, filtering, and amplification. The current–voltage conversion circuit is shown in **Figure 12A**. The resistance, capacitance, and integrated operational amplifier constitute the photodiode transimpedance amplifier. The resistances RV2 and RV3 and capacitance CV2 constitute the bias voltage with a positive power supply, which is connected to the positive phase input end of the integrated operational amplifier to avoid circuit instability resulting from the saturation of the negative power supply rail when there is no current. The filter circuit is shown in **Figure 12B**. The active filter circuit online design tool Filter Design Tool of the Texas Instruments Company was used to implement circuit design (Duff and Kalb, 2015).

Software System Design

The software system is composed of two modules: initialization and functional control. The initialization module is used for microprocessor initialization, pulse width modulation (PWM) signal initialization, digital-to-analog conversion initialization, keyboard input, and LCD initialization. The functional control module is used for light source time-sharing drive, diffuse reflectance spectrum collection, diffuse reflectance spectrum correction, grain quality detection, keyboard input control, LCD output, and audible output. Light source time-sharing driving means that the digital input/output ports of the microprocessor generate PWM signal and light source control signals, and then the PWM signals make the LEDs generate the modulated light signals. Diffuse reflectance spectrum collection is the collection of modulated light signals through a spectrum collection circuit. Diffuse reflectance spectrum correction is performed by correcting the luminous state of the LEDs in real-time through the correction sample cell (**Figure 13**). Grain quality detection is the coupling of diffuse reflectance spectrum and grain quality detection model to calculate quality components in real-time. And, the grain quality detection model is converted into the quality detection library file in c format through MATLAB coder, and the quality detection library file is called in the microprocessor program to realize real-time detection.

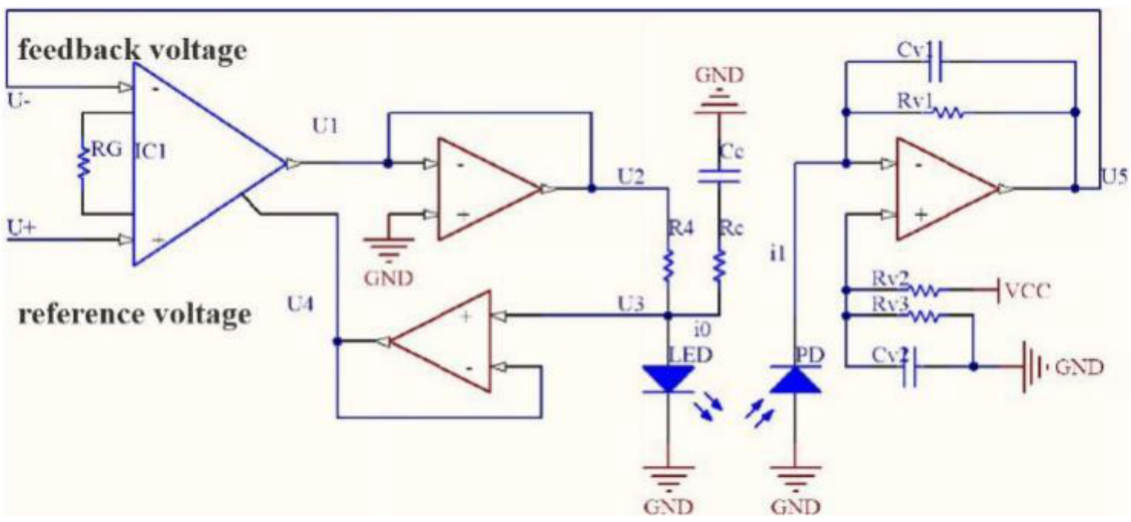


FIGURE 11 | Optical power feedback drive circuit.

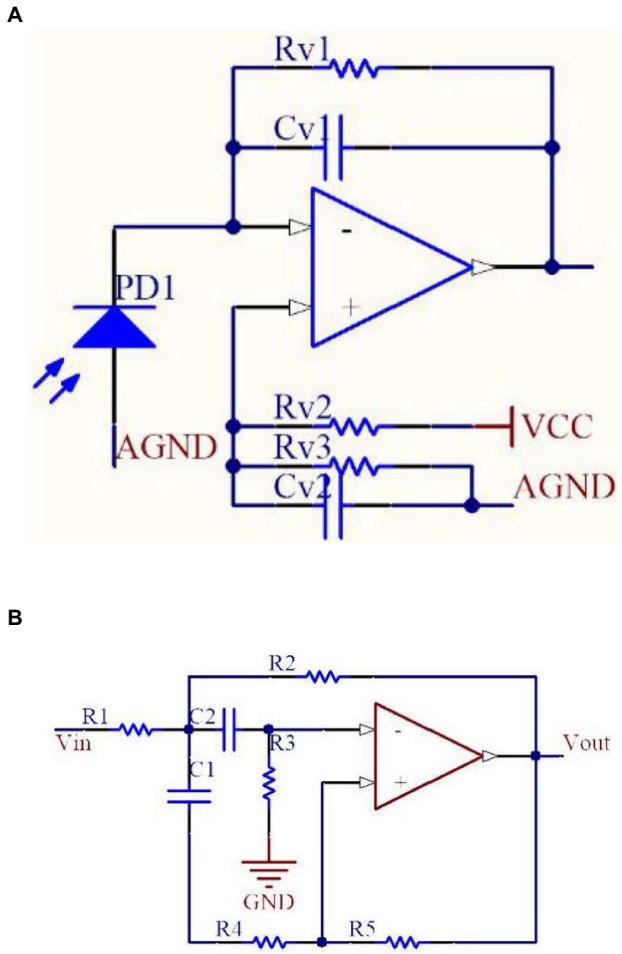


FIGURE 12 | (A) Current–voltage conversion circuit, **(B)** MFB-type second-order band-pass filter circuit.

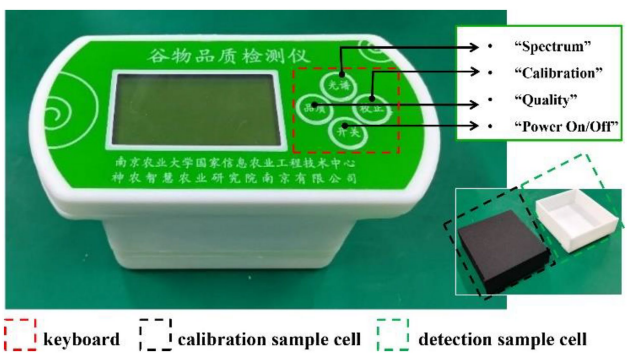


FIGURE 13 | The physical image of the NIR PSPMWG.

As shown in **Figure 13**, the NIR PSPMWG has three functions: "correction," "spectrum" and "quality." In the "correction" mode, the function of correcting the diffuse reflectance spectrum is realized. In the "spectrum" mode, the light source is first driven by time-sharing, and then the diffuse reflectance spectrum is acquired in real-time through the diffuse reflectance spectrum collection; In the "quality" mode, the grain quality information is generated in real-time through the grain quality detection model, and then displayed through the LCD and output through speech synthesis.

TESTS AND RESULTS

Tests

Three types of tests were conducted: correction, performance, and calibration. In the correction test, multiple PSPMWGs were used to measure the relationship between the RDR and the standard diffuse reflectance of the diffusion fabric. The purpose is to ensure the universal application of the grain quality detection model in multiple PSPMWGs. Diffusion fabrics with standard diffuse reflectance of 6.5, 25, and 48% were used as the test material in the correction test. The RDR of the diffusion fabrics was measured by the PSPMWG, and the mean value of each diffusion fabric measured five times was taken as its RDR. Then, the linear relationship between the RDR and the standard diffuse reflectance of the diffusion fabric was established.

In the performance test, the RDR of the same grain was measured several times to verify the stability of the NIR PSPMWG. Two wheat grain samples (sample 1: Yangmai 23, sample 2: Ningmai 13) were selected in the performance test. The PSPMWG was used to collect the noise level of the RDR instrument five times, and the repeatability performance of the RDR instrument was accessed under the repeated sample condition.

In the calibration test, a relational model between the RDR and the physicochemical analyses values was built such that the moisture and protein contents could be directly obtained once the RDR was measured by the PSPMWG. Twenty-four types of wheat grain samples (a total of 48 samples, each variety had two grains, and each sample weighed 250 g) were selected in the calibration test. The PSPMWG was used to collect the RDR of wheat grain; the mean value of each sample measured five times was taken as the RDR, and the moisture and protein contents of 48 samples were measured *via* physicochemical analyses. Then, the relational model between RDR and physicochemical analyses values was created using the chemometrics method, i.e., the wheat grain quality detection model.

Materials

The test equipment includes PSPMWG and diffusion fabrics. A physical image of the NIR PSPMWG is shown in **Figure 13**. The phenotypic sensor can obtain the RDR of four wavelengths in real-time, i.e., 780, 910, 980, and 1,050 nm. The diffuse reflectance of the three diffusion fabrics was 6.5, 25, and 48%. A physical image is shown in **Figure 14**.

Determination of Quality Parameters

In the physicochemical analyses of moisture and protein contents of wheat, the conventional ISO method (ISO, 2010) was used

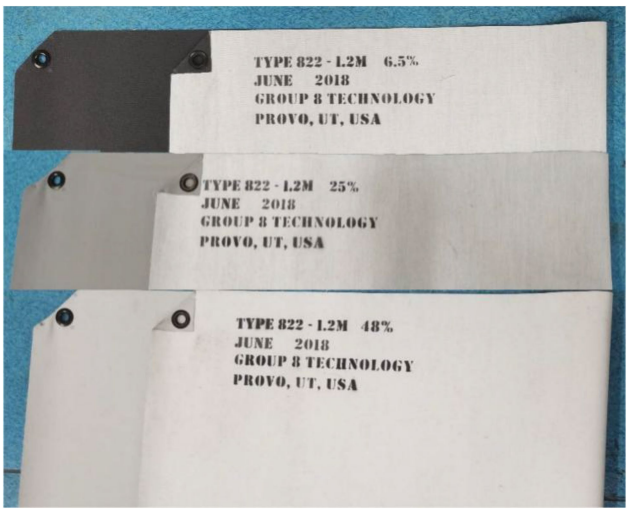


FIGURE 14 | Three diffusion fabrics.

to determine the moisture of wheat. Specifically, the sample was weighed on a dried aluminum container with constant weight and dried at (130 ± 3) °C for 90 min. Then, the sample was taken out and cooled to air temperature in the dryer, which was weighed again. The protein content of wheat was determined by the semi-micro Kjeldahl method (ISO International Standard, 2006), and the protein content was calculated by multiplying the nitrogen content by 5.7. The moisture and protein contents of each sample were measured twice, and the mean value was taken as the final data.

Modeling Methods

There was a nonlinear relationship between the RDR of wheat grain and its physicochemical analyses values. Due to the interaction between the moisture of wheat grain and the components such as protein, the noise of the sensor and other factors can also cause nonlinearity, and the ideal model cannot be obtained by the linear correction method. The artificial neural network method has high nonlinear expression ability and is widely used to establish nonlinear NIR analysis models. In the PSPMWG, the Neural Fitting (nftool) app in MATLAB was used to establish the grain quality, detection model. The nftool is a two-layer feed-forward network with sigmoid hidden neurons and linear output neurons, that can fit multi-dimensional mapping problems arbitrarily well， neural network is shown in **Figure 15**. 48 samples were randomly divided into a Training set (28 samples), Validation set (10 samples), and Test set (10 samples). In the nftool, the Training set is presented to the network during training, and the network is adjusted according to its error; the Validation set is used to measure network generalization, and to halt training when generalization stops improving, and the Test set provides an independent measure of network performance during and after training. The RDR of wheat grain was used as the input, and the moisture and protein content were used as the output. The model was evaluated using the coefficient of determination (R^2), RMSE, and MAE.

(1) (R2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{i\_true} - y_{i\_predict}\right)^2}{\sum_{i=1}^{n}\left(y_{i\_true} - \overline{y}_{true}\right)^2} \quad (4)$$

where y_{i_true} is the physicochemical value of sample i, $y_{i\_predict}$ is the predicted value of sample i, $\overline{y}_{true}$ is the average physicochemical value of all samples, and n is the number of samples. With the same real value range, the closer R^2 is to 1, the better the regression or prediction of the model.

(2) RMSE

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{i\_true} - y_{i\_predict}\right)2}{n}} \quad (5)$$

where y_{i_true} is the physicochemical value of sample i, $y_{i\_predict}$ is the prediction value of sample i, and n is the number of samples. The smaller RMSE, the better predictability of the model (Hussain et al., 2019).

(3) MAE

$$MAE = \frac{\sum_{i=1}^{n}\left|y_{i\_true} - y_{i\_predict}\right|}{n} \quad (6)$$

where y_{i_true} is the physicochemical value of sample i, $y_{i\_predict}$ is the prediction value of sample i, and n is the number of samples. The smaller MAE, the better predictability of the model. MAE can reflect the actual situation of the predicted value error. The smaller MAE, the smaller the prediction error.

Results and Discussion

Correction Test

The diffuse reflectance correction of the NIR PSPMWG is shown in **Figure 16**. Here, 780, 910, 980, and 1,050 refer to the diffuse reflectance corresponding to the wavelengths of 780, 910, 980, and 1,050 nm, respectively. The RDR of the diffuse fabrics collected by the phenotypic sensor was used as the independent variable and the standard reflectance of the diffuse fabrics was used as the dependent variable; then, the diffuse reflectance correction equation was established. The values of R^2 were all above 0.99. The diffuse reflectance correction equations of 780, 910, 980, and 1,050 nm were $y = 0.2560 \times -0.1227$, $y = 0.1515 \times -0.0541$, $y = 0.1911 \times -0.0845$, and $y = 0.196 \times -0.1042$.

Performance Test

The noise level and repeatability of the NIR PSPMWG are shown in **Figure 17**. Here, 780–1, 910–1, 980–1, and 1,050–1 refer to the RDR of sample #1 at the wavelengths of 780, 910, 980, and 1,050 nm, respectively. 780–2, 910–2, 980–2, and 1,050–2 are the RDR of sample #2 for the corresponding series of wavelengths.

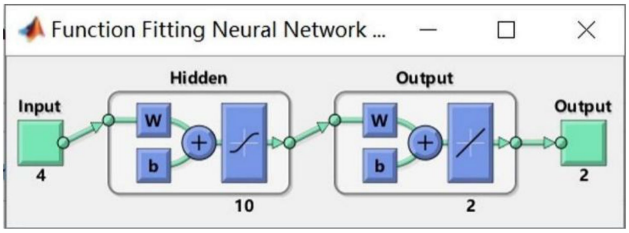


FIGURE 15 | Neural network diagram.

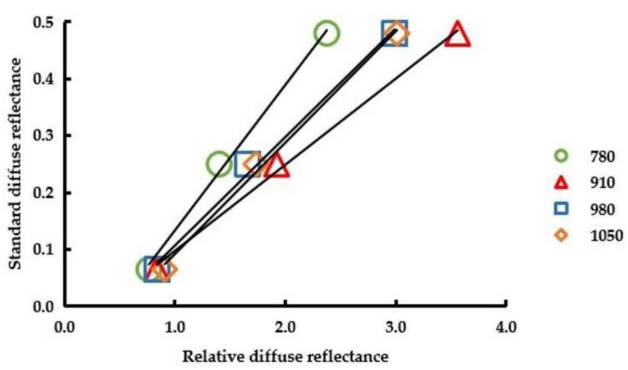


FIGURE 16 | Diffuse reflectance correction of the PSPMWG.

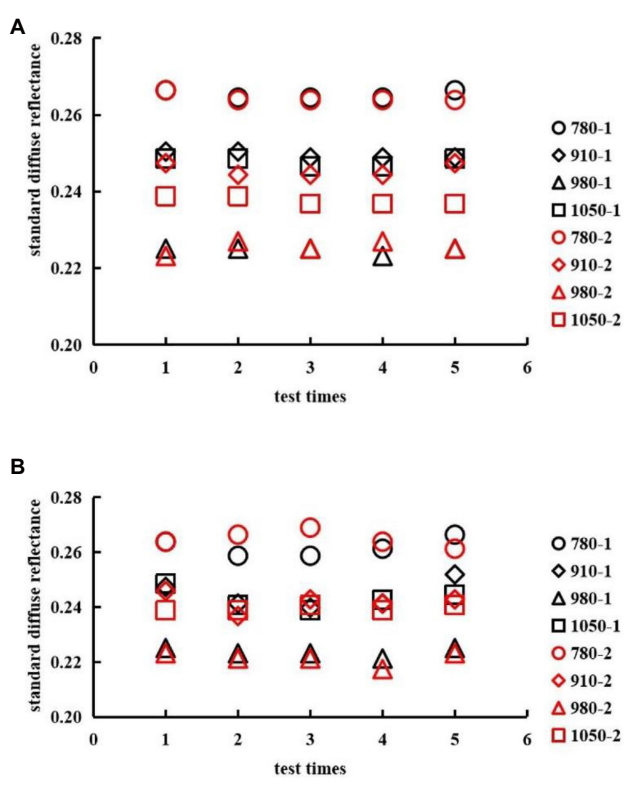


FIGURE 17 | Performance test of the PSPMWG. **(A)** Noise level, **(B)** Repeatability.

In the noise level test, the standard deviations of RDR of sample #1 were 0.11, 0.08, 0.09, and 0.11%. The standard deviations of RDR of sample #2 were 0.11, 0.17, 0.16, and 0.11%. In the repeatability test, the standard deviations of RDR of sample #1 were 0.33, 0.51, 0.16, and 0.38%. The standard deviations of RDR of sample #2 were 0.29, 0.33, 0.23, and 0.11%. In summary, the NIR PSPMWG can accurately collect the RDR of wheat grains.

Calibration Test

The relationship between the detection values obtained by the NIR PSPMWG through the quality detection model and the physicochemical values is shown in **Figure 18**. In the figure, Target represents the physicochemical values of protein and moisture, Output represents the detection values. In the prediction of protein, the RMSE of the Training set was 0.7981% and MAE was 0.6060%; the RMSE of Validation set was 0.4218% and MAE% was 0.6229; the RMSE of the Test set was 0.4866% and MAE was 0.6515%. In the prediction of moisture, the RMSE of the Training set was 0.4598% and MAE was 0.3634%; the RMSE of Validation set was 0.4295% and MAE was 0.6468%; the RMSE of the Test set was 0.2161% and MAE was 0.3078%. Thus, the established quality detection model can effectively detect the moisture and protein content of wheat grains. The NIR PSPMWG can obtain the moisture and protein content of wheat grains in real-time.

DISCUSSION

The NIR PSPMWG based on NIR spectroscopy can quickly, efficiently, and nondestructively obtain grain quality information. Currently, full-band NIR commercial instruments such as NIR Systems 5,000 spectrometer and FOSS Infratec 1,241 spectrometer have achieved good results in the application of grain quality detection (Bao et al., 2010; Hexiao et al., 2011; Lin et al., 2014; Chen et al., 2019). Liu et al. (Hexiao et al., 2011) established a neural network model for wheat protein detection. The MSE of the calculated values and actual values of wheat protein content ranged from 0.799 to 0.09, and the average MSE of 0.178 transformed into RMSE of 0.4219%. This result is similar to the detection effect of our phenotypic sensors. However, these instruments are complex, bulky, non-portable, and expensive. Moreover, since only spectrum information was collected, and there was no quality detection model, these devices cannot be used for real-time on-site detection of grain quality. Given the above problems, in this study, we developed a NIR PSPMWG. Compared with existing full-band commercial instruments, the phenotypic sensor uses four wavelengths that are sensitive to grain quality and an optical system with a multi-source circular structure. The dimensions of the optical system were greatly reduced while ensuring that the diffuse reflectance spectrum of multi-grain wheat was collected. In existing commercial spectrometers, one needs to customize the function or manually calculate the quality information through the collected spectral information to realize grain quality detection the customization is costly and the manual calculation leads to poor real-time performance; thus, it is

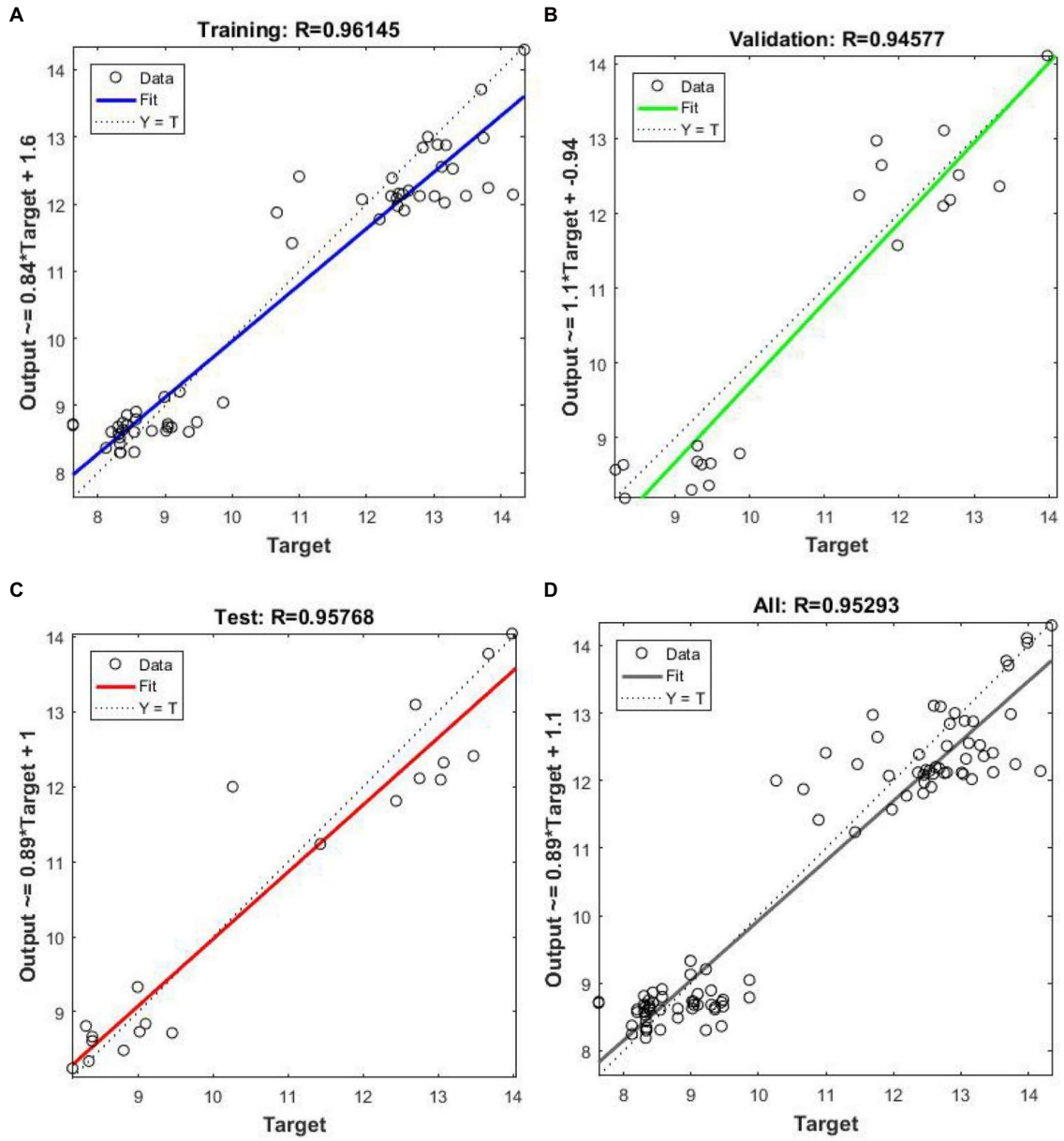


FIGURE 18 | Scatter plot of the detection values and the physicochemical values. **(A)** Training set, **(B)** Validation set, **(C)** Test set, **(D)** All values.

difficult to implement these instruments in actual applications. However, by collecting the RDR of wheat grain using the designed sensor, and coupling it with the quality detection model, the grain quality information was obtained in real-time.

Due to the different sensitive wavelengths of different grain quality components, the PSPMWG needs to change the wavelength of its light source for specific detection objects. In addition, the shape, size, and size distribution of different varieties of wheat grains affect the scattering coefficient, which can lead to measurement error in the RDR. Further studies are needed to explore the methods that can reduce the difference of scattering coefficient and improve the accuracy of RDR and ultimately the detection accuracy of the PSPMWG.

CONCLUSION

(1) A NIR PSPMWG was developed. An optical system with a multi-source circular structure was used to avoid the difference of incident light spots of multiple light sources. The light source drive circuit based on optical power feedback and the weak optical signal conditioning spectral acquisition circuit was designed to improve the luminous stability and signal-to-noise ratio of the sensor. The standard deviation of RDR at four wavelengths was less than 0.17 and 0.51% in noise level and repeatability test, indicating that the sensor developed in this study can accurately collect RDR of wheat grain.

(2) The quality detection model of wheat grains was established. The RDR of wheat grains was collected by the phenotypic sensor, and the physicochemical values of moisture and protein were obtained by physicochemical analyses. In the Test set, the RMSEs of protein and moisture content of wheat were 0.4866 and 0.2161%, the MAEs were 0.6515 and 0.3078%, indicating that the phenotypic sensor can achieve real-time, on-site acquisition of RDR and quality information of wheat grains.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YL: data curation, visualization, writing—original draft. YL and DL: formal analysis. JN: funding acquisition and project administration. YL, DL, and HL: investigation. JN and YL: methodology and writing—review and editing. WC and YZ: resources. YL and XJ: software and supervision. XJ, YL, and HL: validation. All authors have read and agreed to the published version of 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 [grant number 31871524]; the Primary Research & Development Plan of Jiangsu Province of China [grant numbers BE2021308, BE2021304]; Six Talent Peaks Project in Jiangsu Province [grant number XYDXX-049]; and the 111 Project [grant number B16026].

ACKNOWLEDGMENTS

We would like to thank all the researchers in the Intelligent Equipment Research Group of the National Engineering and Technology Center for Information Agriculture and all the foundations for this research.

REFERENCES

Arazuri, S., Ignacio Arana, J., Arias, N., Arregui, L. M., Gonzalez-Torralba, J., and Jaren, C. (2012). Rheological parameters determination using near infrared technology in whole wheat grain. *J. Food Eng.* 111, 115–121. doi: 10.1016/j.jfoodeng.2012.01.017

Bao, J., Cai, Y., and Corke, H. (2010). Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy. *J. Food Sci.* 66, 936–939. doi: 10.1111/j.1365-2621.2001.tb08215.x

Bec, K., Grabska, J., and Huck, C. W. (2020). Principles and applications of miniaturized near-infrared (NIR) spectrometers. *Chem Eur J.* 27, 1514–1532. doi: 10.1002/chem.202002838

Caporaso, N., Whitworth, M., and Fisk, I. D. (2018). Near-infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Appl. Spectrosc. Rev.* 53, 667–687. doi: 10.1080/05704928.2018.1425214

Chen, J., Li, M., Pan, T., Pang, L., Yao, L., and Zhang, J. (2019). Rapid and non-destructive analysis for the identification of multi-grain rice seeds with near-infrared spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 219, 179–185. doi: 10.1016/j.saa.2019.03.105

Delwiche, S. R. (1995). Single wheat kernel analysis by near-infrared transmittance: protein content. *Cereal Chem.* 72, 112–114. doi: 10.1021/bp00031a017

Delwiche, S. R., McKenzie, K. S., and Webb, B. D. (1996). Quality characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chem.* 12, 286–288. doi: 10.1021/bp960006l

Duff, M. N., and Kalb, A. J. (2015). US, assignee. Filter design tool. the United States patent US 8,930,874 B2.

El-Mesery, H. S., Mao, H., and Abomohra, A. E.-F. (2019). Applications of non-destructive technologies for agricultural and food products quality inspection. *Sensors* 19:846. doi: 10.3390/s19040846

Hexiao, L., Mingliang, L., and Laijun, S. (2011). Quality analysis of wheat based on BP neural network and Near Infrared Reflectance Spectroscopy. presented at: 2011 International Conference on New Technology of Agricultural Engineering [v1]; Zibo, China. Available at: https://ieeexplore.ieee.org/document/5943907

Hidaka, Y., Kurihara, E., Hayashi, K., Noda, T., and Sashida, K. (2011). Near-infrared spectrometer for a head-feeding combine for measuring rice protein content. *Jpn. Agric. Res. Q.* 45, 63–68. doi: 10.6090/jarq.45.63

Hussain, N., Sun, D.-W., and Pu, H. (2019). Classical and emerging non-destructive technologies for safety and quality evaluation of cereals: a review of recent applications. *Trends Food Sci. Technol.* 91, 598–608. doi: 10.1016/j.tifs.2019.07.018

ISO International Standard (2006) Cereals and pulses. Determination of the nitrogen content and calculation of the crude protein content. Kjeldahl method.

ISO (2010). Cereals and cereal products – determination of moisture content – reference method (ISO 712-2009).

Li, R., Kawamura, S., Fujita, H., and Fujikawa, S. (2013). Near-infrared spectroscopy for determining grain constituent contents at grain elevators. *Eng. Agric. Environ. Food* 6, 20–26. doi: 10.1016/S1881-8366(13)80013-4

Lin, C., Chen, X., Jian, L., Shi, C., Jin, X., and Zhang, G. (2014). Determination of grain protein content by near-infrared spectrometry and multivariate calibration in barley. *Food Chem.* 162, 10–15. doi: 10.1016/j.foodchem.2014.04.056

Liu, Y., Li, Y., Peng, Y., Han, D., Ding, J., and Wang, Q. (2019). Portable rapid nondestructive detecting instrument for multi-quality parameters of Rice. *Trans. Chin. Soc. Agric. Mach.* 50, 358–364. doi: 10.6041/j.issn.1000-1298.2019.08.038

Priya, B. S., Kumaravelu, C., Gopal, A., and Stanley, P. (2015). Classification of rice varieties using near-infra red spectroscopy. presented at: 2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR); 10–12 July 2015. Available at: https://ieeexplore.ieee.org/document/7358524

Risius, H., Hahn, J., Huth, M., Tölle, R., and Korte, H. (2014). In-line estimation of falling number using near-infrared diffuse reflectance spectroscopy on a combine harvester. *Precis. Agric.* 16, 261–274. doi: 10.1007/s11119-014-9374-5

Santos, C. D., Lopo, M., Páscoa, R., and Lopes, J. A. (2013). A review on the applications of portable near-infrared spectrometers in the agro-food industry. *Appl. Spectrosc.* 67, 1215–1233. doi: 10.1366/13-07228

The National Health and Family Planning Commission of the People's Republic of China (2016). National Food Safety Standard Determination of Moisture in Foods. China: The National Health and Family Planning Commission of the People's Republic of China.

The National Health and Family Planning Commission of the People's Republic of China & State Food and Drug Administration (2010). National food safety standard determination of protein in foods. China: The National Health and family planning Commission of the People's republic of China & State Food and Drug Administration.

Wang, S. (2010). Infrared spectroscopy for food quality analysis and control. *Trends Food Sci. Technol.* 21:52. doi: 10.1016/j.tifs.2009.08.004

Wen, M., and Ji, H. (2004). Development of portable LED-based NIR integrity wheat component measuring apparatus. *Spectrosc. Spectr. Anal.* 24, 1276–1279. doi: 10.3321/j.issn:1000-0593.2004.10.035

Wu, T., Armstrong, P. R., Zhang, H., Yang, L., and Gao, W. (2018). Investigation on individual wheat kernel quality prediction device with stereoscopic light source. *Trans. Chin. Soc. Agric. Mach.* 49, 363–369. doi: 10.6041/j.issn.1000-1298.2018.10.041

Wu, H., and Yang, F., Zhou, L. (2018). The design and analysis based on the Fresnel condenser lens. presented at: 2nd International Forum on Management, Education and Information Technology Application (IFMEITA 2017).

Yang, H., Xu, L., Chen, K., He, Q. S., He, S. R., Tan, Q. F., et al. (2005). The effect of the photoelectric detector on the accuracy of the spectrometer. *Spectrosc. Spectr. Anal.* 25, 1520–1523. doi: 10.3321/j.issn:1000-0593.2005.09.039

Yao, L., Wu, R., Wu, S., Jiang, X., Zhu, Y., Cao, W., et al. (2020). Design and testing of an active light source apparatus for crop growth monitoring and diagnosis. *IEEE Access.* 8, 206474–206490. doi: 10.1109/access.2020.3037966

Zhang, Y., Zhao, B., Wang, H., Mao, W., and Zhang, X. (2009). Technology of near infrared spectroscopy analysis system design based on fixed grating and charge-coupled device. *Trans. Chin. Soc. Agric. Eng.* 25, 176–181. doi: 10.3969/j.issn.1002-6819.2009.09.031

Zhu, L., Ma, W., Hu, J., Zheng, Y. Y., Tian, Y. X., Guan, Y. J., et al. (2015). Advances of NIR spectroscopy technology applied in seed quality detection. *Spectrosc. Spectr. Anal.* 35, 346–349. doi: 10.3964/j.issn.1000-0593(2015)02-0346-04

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fpls.2022.872555

Automatic Detection and Counting of Wheat Spikelet Using Semi-Automatic Labeling and Deep Learning

Ruicheng Qiu[1,2], *Yong He*[1,2]* *and Man Zhang*[3]

[1] College of Biosystem Engineering and Food Science, Zhejiang University, Hangzhou, China, [2] Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Zhejiang University, Hangzhou, China, [3] Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing, China

The number of wheat spikelets is an important phenotypic trait and can be used to assess the grain yield of the wheat crop. However, manual counting of spikelets is time-consuming and labor-intensive. To develop a cost-effective and highly efficient phenotyping system for counting the number of spikelets under laboratory conditions, methods based on imaging processing techniques and deep learning were proposed to accurately detect and count spikelets from color images of wheat spikes captured at the grain filling stage. An unsupervised learning-based method was first developed to automatically detect and label spikelets from spike color images and build the datasets for the model training. Based on the constructed datasets, a deep convolutional neural network model was retrained using transfer learning to detect the spikelets. Testing results showed that the root mean squared errors, relative root mean squared errors, and the coefficients of determination between the automatic and manual counted spikelets for four wheat lines were 0.62, 0.58, 0.54, and 0.77; 3.96, 3.73, 3.34, and 4.94%; and 0.73, 0.78, 0.84, and 0.67, respectively. We demonstrated that the proposed methods can effectively estimate the number of wheat spikelets, which improves the counting efficiency of wheat spikelets and contributes to the analysis of the developmental characteristics of wheat spikes.

Keywords: wheat spikelet, spike, annotation, deep learning, computer vision

OPEN ACCESS

Edited by:
Yongliang Qiao,
The University of Sydney, Australia

Reviewed by:
Karansher Singh Sandhu,
Bayer Crop Science, United States
Yong Suk Chung,
Jeju National University, South Korea
Xin Zhang,
Mississippi State University, United States

***Correspondence:**
Yong He
yhe@zju.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *09 February 2022*
Accepted: *21 April 2022*
Published: *30 May 2022*

Citation:
Qiu R, He Y and Zhang M (2022) Automatic Detection and Counting of Wheat Spikelet Using Semi-Automatic Labeling and Deep Learning. Front. Plant Sci. 13:872555. doi: 10.3389/fpls.2022.872555

INTRODUCTION

Breeding of high-yield wheat (*Triticum aestivum* L.) cultivars is crucial for ensuring food safety, as wheat is a staple food in the world. Researchers have reported that wheat yield is highly associated with several phenotypic traits, such as spike number per unit area (SNA), grain number per spike (GNS), and thousand-grain weights. It is broadly agreed that improving the SNA and GNS of wheat is important to increase the wheat yield (Vahamidis et al., 2019). A wheat spike consists of many spikelets and a rachis, and each spikelet contains two or more florets. In general, only 1–3 florets can become fertile florets and develop into grains. Improvements in spikelet and floret (floret primordia and fertile floret) numbers contribute significantly to an increment in GNS (García et al., 2014). In addition, the number of spikelets, fertile florets, and grains would enable the calculation of the spikelet fertility, fertile floret proportion, and grain/fertile floret ratio to further assess the

spike characteristics (Guo et al., 2018). Therefore, counting the number of spikes, spikelets, and florets during the breeding process is of great importance for screening high-yield wheat cultivars. However, conventional methods for the manual phenotyping of wheat spike traits are time-consuming and labor-intensive, which in turn delays progress in breeding programs. Consequently, it is urgent to develop an efficient method to accurately and quickly acquire phenotypic traits of wheat spikes. Recent advances in computer vision technology provide innovative ways to assess phenotypic traits of wheat spikes, and techniques like color, X-ray, and computed tomography (CT) imaging have been investigated.

Color imaging has been widely applied to measure the phenotypic traits of crops (Qiu et al., 2018). With respect to wheat spikes, many researchers focused on the automatic detection and counting of spikes. The color features (e.g., red, green, and blue), texture features (e.g., gray level co-occurrence matrix), and image features (e.g., contour and edge) were selected or combined to train a model (e.g., support vector machine and neural network models) using supervised learning, to facilitate the detection of wheat spikes (Li et al., 2017; Zhou et al., 2018; Xu et al., 2020). In addition, Genaev et al. (2019) analyzed wheat spike images in the laboratory and estimated the morphometric traits of spikes, such as spike length, width, and circularity. Liu et al. (2017), Kaya and Saritas (2019) developed a real-time sorting system and an application program, respectively, to identify each grain and count the number of wheat grains. However, color imaging has not been adequately exploited for the detection and counting of wheat spikelets. Researchers have also explored the usefulness of deep learning techniques for measuring spike phenotypic traits. Concerning target detection, deep convolutional neural network (DCNN) models have been widely implemented. The Faster Region-based Convolutional Network (RCNN) model (Madec et al., 2019) and Mask RCNN model (Qiu et al., 2019) were retrained to detect wheat spikes from color images captured in field conditions. Pound et al. (2017) developed a DCNN model and presented the Annotated Crop Image Dataset (ACID) to count wheat spikes and spikelets. Besides, Khoroshevsky et al. (2021) developed a deep neural network to detect and count the number of spikelets per spike in a field. Chen et al. (2021) proposed a method to train deep networks on data with reduced numbers of annotations to count wheat spikelets. TasselNetv2 (Xiong et al., 2019b), SpikeletFCN (Alkhudaydi et al., 2019), SpikeSegNet (Misra et al., 2020), and DeepCount (Sadeghi-Tehran et al., 2019) were also developed to detect and count wheat spikes or spikelets in the field and laboratory. These studies revealed that deep learning techniques are promising for detecting and counting wheat spikes and spikelets. However, one of the current challenges is to obtain a large number of labeled datasets to train the deep learning models. Manual labeling is a heavy burden. Furthermore, some researchers applied adversarial learning to leaf and spikelet countings with unsupervised training (Giuffrida et al., 2019; Hu et al., 2019; Ayalew et al., 2020), but the models are difficult to train.

Furthermore, X-ray and CT imaging have been explored to non-destructively measure phenotypic traits of wheat spikes. Both X-ray and CT imaging can measure the inner structures of spikes and acquire information on grains (Duan et al., 2011; Xiong et al., 2019a; Yu et al., 2021), which can be further used to distinguish and count the filled spikelets (Zhou et al., 2021). As for CT imaging, it can reconstruct spikes in three dimensions as well. Thus, wheat spike and grain traits, including spike height, grain number, grain width, height, and depth, can be extracted and measured (Hughes et al., 2017; Xiong et al., 2019a). Although X-ray and CT imaging can provide considerable inner and spatial information about wheat spikes, they are expensive and the imaging systems are complicated, which limits their application.

Currently, color images are easily acquired at a low cost, and most studies focus on the automatic measurement and counting of wheat spikes using color imaging. But studies regarding the automatic detection of wheat spikelets have not been well reported, and the usefulness of color imaging for counting spikelets needs to be further investigated. Deep learning has been proved very helpful for the phenotyping of wheat spikes. However, manual annotation is a laborious and tedious process. Therefore, this study focuses on detecting and counting spikelets using color imaging and deep learning techniques. Four common wheat lines were selected as the objects of this study, and the color images of their spikes were collected in the laboratory. The objectives of this study were to (1) develop an unsupervised learning-based method to automatically detect and label spikelets from spike color images and build the datasets for DCNN model training and (2) train a DCNN model that can detect and count spikelets. The counting results of spikelets will be compared with manual countings to evaluate the performance of the proposed methods.

MATERIALS AND METHODS

The counting system mainly included the following steps: image collection, image annotation, spikelet detection and counting, and performance evaluation. Several algorithms were used and developed to detect spikelets and build datasets in the section of image annotation, and a DCNN model was trained using a deep learning method to detect spikelets in the section of spikelet detection and counting. Each step is described in detail in the following sections. A desktop computer with AMD R5-2600 CPU, NVIDIA GeForce GTX 1070, 8G RAM, and Windows 10 64-bit system was utilized to process the images of wheat spikes and test the proposed methods. Image annotation and spikelet detection and counting were conducted using Matlab R2021b software and Tensorflow, respectively.

Image Collection

The experiment was conducted in a field station in the city of Hengshui (38°21′ N, 115°65′ E), Hebei Province, China. The cultivated wheat lines were Shiyou 20, Shannong 25, Liangxing 99, and Shenmai 818, which were largely cultivated in the North China Plain. Their wheat spikes with awns have different shapes and colors. Wheat spike samples for each wheat line were randomly selected and collected on May 18, 2021, and most of the wheat spikes were at the grain filling stage of development. At this stage, the wheat spikelets have basically been formed, wheat awns and glumes have not begun to senesce, and the

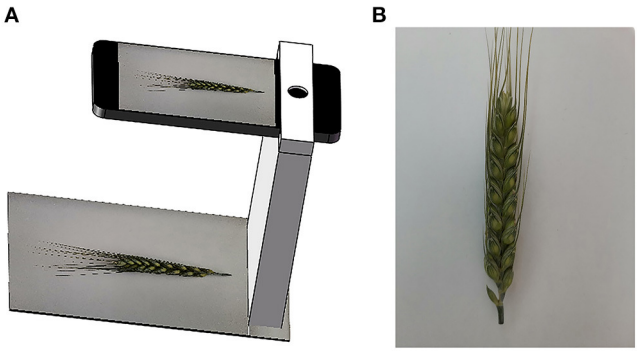


FIGURE 1 | Schematic diagram of wheat spike image capturing system and captured images. **(A)** Image capturing system and **(B)** a spike image. A spike example of Shiyou 20 was used in **Figures 1**–**4** to explain our work.

FIGURE 2 | Schematic diagram of region of interest.

color contrast of spikelet and awn is great (Qiu et al., 2019). The collected wheat spikes were transferred to the laboratory on the same day. Wheat spikes were placed on a flat board, and a HUAWEI Honor 9X PRO smartphone was used to capture spike images, as illustrated in **Figure 1A**. A spike was captured twice by rolling 180° to acquire its images on two sides. The captured spike images (**Figure 1B**) have a 4,000 × 3,000 pixel resolution in a JPEG (Joint Photographic Experts Group) compressed format. In the image, the apical spikelet is located on the top of the spike, other spikelets are located on both sides of the spike rachis, and the glumes of spikelets are clear. Spikelets can be counted by detecting the glumes in the spike image. Finally, more than 300 spike images for each wheat line were collected.

FIGURE 3 | Spike segmentation. **(A)** Gray (S component) image of wheat spike and **(B)** segmentation results of wheat spike.

Image Annotation

In this study, an unsupervised learning method based on the watershed algorithm was developed to annotate the spike images. In addition, a DCNN model was trained to optimize the annotation. The method contains several steps, which are described in detail in the following sections. The proposed labeling method was implemented to process the collected spike images of four wheat lines.

Image Preprocessing

The color images of raw spike have high resolution, which is not conducive to the subsequent image processing. Therefore, a region of interest (ROI) was set to the color images of the wheat spike, to reduce the computation. The ROI is shown in **Figure 2**, and the whole spike was extracted following the steps depicted in the figure. Moreover, the extracted images were downsampled using the bicubic method to reduce image resolutions. The scale factor was set to 0.3, and the spike images were reshaped to 256 × 1,021 pixel resolution.

Spike Extraction

It is of significance to segment spikes from the spike images in advance, to minimize detection areas and improve the detection efficiencies and precision of spikelets. Although the color characteristics of the spike and background in the spike images are different, the color characteristics of spikes for

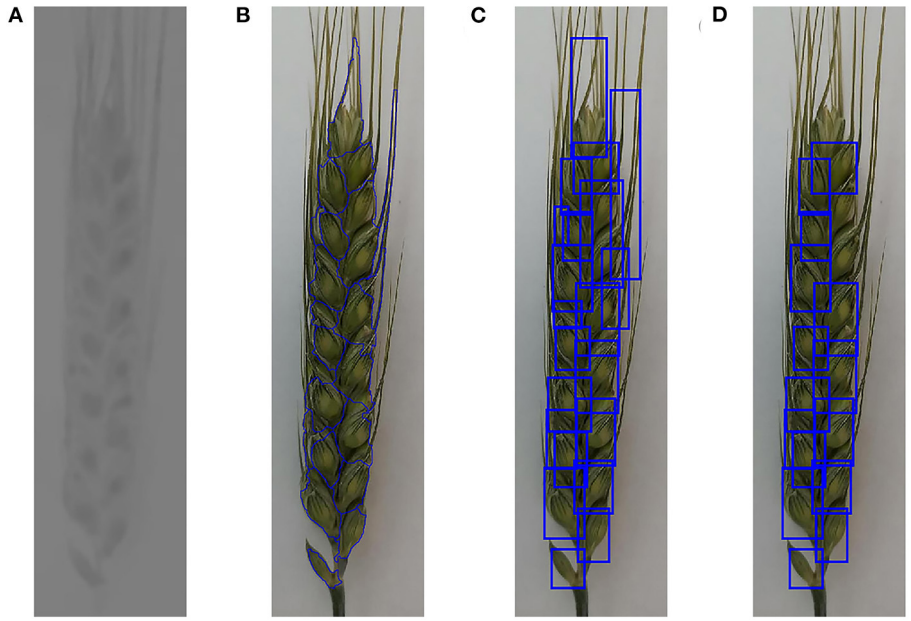


FIGURE 4 | Spikelet annotation. **(A)** Gray (Cb component) image of wheat spike, **(B)** segmentation results of watershed algorithm, **(C)** MER for all candidate spikelet regions, and **(D)** optimized MER for candidate spikelet regions.

TABLE 1 | The number of spikelets labeled by the watershed algorithm for training and validation datasets.

Wheat line	Training	Validation	SUM
Shiyou 20	1,097	267	1,364
Shannong 25	1,142	265	1,407
Liangxing 99	1,095	269	1,364
Shenmai 818	1,304	333	1,637
SUM	4,638	1,134	5,772

different wheat lines are not uniform in RGB (Red, Green, and Blue) color space, which makes it difficult to apply RGB color features to segment and extract spikes. After tests, HSV (Hue, Saturation, and Value) color space was applied to process the reshaped images after image preprocessing. S component of the spike in the color image was calculated using the function "rgb2hsv" provided by Matlab to generate the gray image to represent the color characteristics of the spike, as shown in **Figure 3A**. The contrast between spike and background is stark so that the spike can be detected and extracted accurately. Otsu's algorithm was implemented to process the gray image of the spike and generate its binary image (**Figure 3B**). In the binary image, the pixel values of spike and background are 1 and 0, respectively. As shown in **Figure 3B**, the spike was successfully extracted.

Spikelet Segmentation and Annotation

The color characteristics of spikelets are different from the other parts of spikes, which contributes to the detection and segmentation of spikelets. Specifically, some parts of glumes are prominent in the color images.

The color images of the spike were transformed into several color spaces, such as RGB, HSV, and YCbCr, to find suitable features for spikelet segmentation. Testing results showed that the color features of spikelets and their boundaries were significant in YCbCr color space. Cb component of the spike in the color image was calculated using the function "rgb2ycbcr" provided by Matlab to generate the gray image to represent the color characteristics of spikelets. Then, a bilateral filter was used to enhance the contrast between the spikelets and their boundaries in the gray image, as shown in **Figure 4A**. By performing a per-pixel dot product between **Figures 4A**, **3B**, the spike gray image was extracted, and the gray values of ground pixels were set to 0.

In the spike gray image, the values of glumes are lower than that of their surrounding pixels. Based on this characteristic, the watershed algorithm can be used to segment the spikelets from spikes. However, the watershed algorithm usually produces over-segmentation. To improve the segmentation accuracy, a watershed algorithm, marked by a local minimum threshold, was applied to process the extracted gray image of the spike. Tests showed that when the local minimum threshold was set to 3, the obtained results can remove some local minimums and avoid over-segmentation. The segmented boundaries generated by the watershed algorithm can divide the spike into many regions (**Figure 4B**), which contained many candidate spikelets. After the initial segmentation of the spikelet, the areas and minimum enclosed rectangles (MER) for all the candidate spikelet regions were calculated (**Figure 4C**). According to the sizes and shapes of spikelets, the regions whose areas were between 1,000 and 5,000

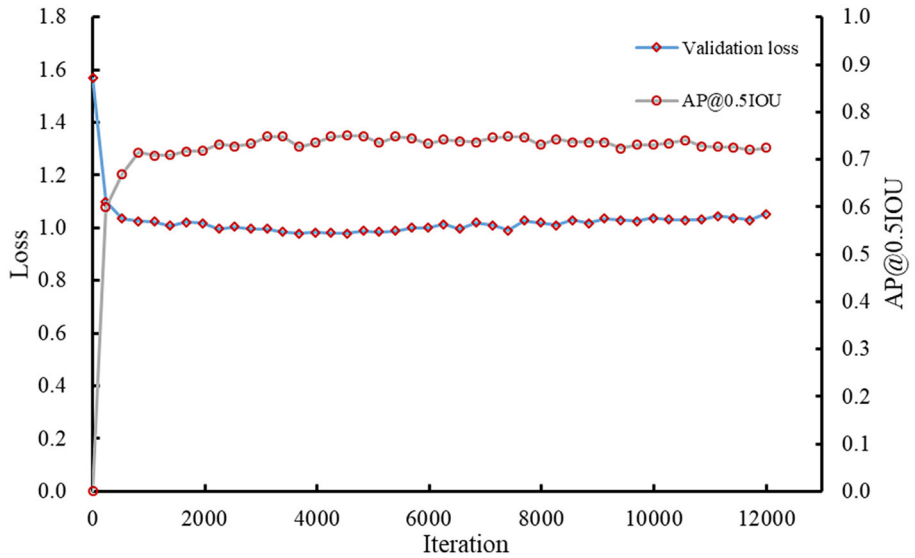


FIGURE 5 | Loss and AP@0.5IOU of the Faster RCNN model for validation.

TABLE 2 | The number of labeled spikelets for training and validation datasets.

Wheat line	Labeled by the DCNN model			After manual correction		
	Training	Validation	SUM	Training	Validation	SUM
Shiyou 20	982	224	1,206	1,272	297	1,569
Shannong 25	986	232	1,218	1,238	293	1,531
Liangxing 99	954	227	1,181	1,263	312	1,575
Shenmai 818	1,051	272	1,323	1,251	302	1,553
SUM	3,973	955	4,928	5,024	1,204	6,228

and ratios of length to width for the corresponding MER were in the range of 1–2.5 were reserved, as displayed in **Figure 4D**. After that, the upper-left coordinates, widths, and lengths of MER were saved to obtain the bounding boxes for the initial labeling of spikelets.

Three hundred spike images of each wheat line were randomly selected from their captured images and initially labeled using the described method. As shown in **Figure 4D**, there were some mislabeled or unlabeled spikelets. To acquire better labels, the labeling results were manually checked, and 100 spike images of each wheat line annotated with high accuracy were selected to generate the dataset. XML files for all labeled spike images were generated based on the coordinates of MER, which were used for the subsequent dataset construction and DCNN model training.

DCNN Model Training

In recent years, some DCNN models were developed and widely implemented for object detection. In the present study, Faster RCNN, proposed by Ren et al. (2017), was selected to detect the spikelets given its high detection accuracy, which can classify objects and realize semantic segmentation of spikelets. A Faster RCNN has two main parts, a regional proposal network (RPN) and a Fast RCNN. The RPN is a type of fully convolutional network and generates many anchor regions as candidate bounding boxes. Each anchor is assessed and scored based on its intersection over union (IOU) ratio with the ground truth. The anchors are classified as positive and negative using a Softmax function, and the bounding box regressions of positive anchors are conducted to obtain corrected region proposals, which are used by Fast RCNN for object detection training. In addition, some proposals may overlap with each other, and non-maximum suppression (NMS) is adopted to reduce the number of proposals. The loss function (L) was defined as a function (1).

$$L = L_{rpn\_cls} + L_{rpn\_reg} + L_{rcnn\_cls} + L_{rcnn\_reg} \quad (1)$$

where L_{rpn_cls} and L_{rpn_reg}, and L_{rcnn_cls} and L_{rcnn_reg} represent the classification and bounding box regression losses for RPN and Fast RCNN, respectively.

In this study, Faster RCNN was implemented using Tensorflow object detection API (Huang et al., 2017). The model was pre-trained using the COCO dataset. The Inception-V2 model was used to extract features because of its high speed. The scales (i.e., 0.25, 0.5, 1.0, and 2.0) and ratios (i.e., 0.5, 1.0, and 2.0) were set for the anchors. If the values of the IOU ratio with

the labeled bounding box were higher than 0.6, the anchors were considered to contain a wheat spikelet. The batch size was set to 1 because it saves memory and computation time. The IOU threshold for NMS was set to 0.6. The momentum was fixed to 1. The initial learning rate (LR) was 0.0005. After 4,000 and 8,000 iterations, the LR dropped to 0.0003 and 0.0001, respectively. The training results were recorded once every 60 s.

Before training, 80 spike images were randomly selected from the annotated images (100 images) for each wheat line to generate the training dataset, and the remaining 20 spike images were used as the validation dataset. One hundred seventy spike images for each wheat line that were not annotated were used to build the testing dataset.

The model was retrained and fine-tuned using transfer learning on the desktop computer with Tensorflow 1.10.0, Anaconda 3.5.2, CUDA 9.0, and Python 3.6.7.

If the IOU between a predicted bounding box and a labeled bounding box is higher than a set threshold, the predicted bounding box is considered as a true positive (*TP*). Otherwise, it is considered as a false positive (*FP*). If a labeled spikelet cannot be detected, it is considered as a false negative (*FN*). Then, the recall and precision can be calculated by functions (2) and (3).

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$precision = \frac{TP}{TP + FP} \tag{3}$$

The average precision (AP), which is the area under the precision-recall curve, was applied as an indicator to quantify the performance of the trained Faster RCNN model. In this study, the standard IOU threshold value of 0.5 was used, and the AP@0.5IOU was calculated (Madec et al., 2019).

Dataset Optimization

After training, the spike images in the training and validation datasets were processed again using the trained DCNN model to detect and label the spikelets. The coordinates and confidence scores of detected bounding boxes were saved. The high confidence scores (up to 1) indicated that the detected boxes most probably contain spikelets. The bounding boxes of detected spikelets whose confidence scores were higher than 0.75 were reserved, which were used to update the image annotation and the training and validation datasets.

As the performance of the watershed algorithm for spikelet segmentation and labeling in Section Spikelet Segmentation and Annotation was not perfect, the training and validation datasets used for DCNN model training contain many mislabeled or unlabeled spikelets, and the bounding boxes of spikelet samples generated by the trained DCNN model were incomplete. Therefore, manual corrections were conducted by removing mislabeled regions and adding new spikelet labels to the updated training and validation datasets.

Spikelet Detection and Counting

The ultimate goal of this study was to detect and count the wheat spikelets in spike images. The Faster RCNN model was retrained again using the optimized training and validation datasets in Section Dataset Optimization. The batch size was set to 2, and the momentum was adjusted to 0.8, to prevent overfitting. The initial LR was set to 0.0005. After 6,000 and 10,000 iterations, the LR dropped to 0.0001 and 0.00001, respectively. The training results were recorded once every 15 s.

The retrained model was implemented to detect the spikelets in the spike images of the testing dataset. The confidence score threshold of the detected bounding box was set to 0.75. The final results were counted to obtain the spikelet numbers for all wheat lines.

Performance Evaluation

One hundred seventy spike images of each wheat line in the testing dataset were used to assess the performance of the proposed detection and counting methods for wheat spikelets, which were evaluated using several statistical parameters, including the root mean squared error (*RMSE*), relative *RMSE* (*rRMSE*), and the coefficient of determination (R^2), as described in the following equations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - d_i\right)^2} \tag{4}$$

$$rRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{t_i - d_i}{t_i}\right)^2} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(t_i - d_i\right)^2}{\sum_{i=1}^{n}\left(t_i - \bar{t}_i\right)^2} \tag{6}$$

where n indicates the number of testing images, t_i is the manually counted number of spikelets, d_i is the automatically counted number of spikelets, and $\bar{t}_i$ is the mean value of t_i.

RESULTS

Image Annotation Results

Spikelet Segmentation and Annotation Results

Although the bounding boxes generated by the water algorithm contain some non-spikelet areas and their boundaries show errors, they incorporate the main parts of spikelets. Some unsatisfactory MER were removed in Section Spikelet Segmentation and Annotation, so that many spikelets were inevitably skipped and not annotated. A manual check was performed only to estimate the labeling results, and the boundaries of bounding boxes were not adjusted during the manual check process.

After manual check and selection, the labeled spikelet numbers of 100 spike images (80 images for the training dataset and 20 images for the validation dataset) for each wheat line are summarized in **Table 1**. The results indicated that the proposed method can realize the segmentation and detection of spikelets in spike images, even though the color characteristics of wheat lines

are different. Also, it is effective for the initial spikelet labeling of different wheat lines. Finally, a total of 5,772 spikelets were roughly labeled. All the labeled spike images are provided in the **Supplementary Material**.

Model Training and Dataset Optimization Results

The Faster RCNN model was trained using the initial labeled spikelet samples. As presented in **Table 1**, the training and validation datasets contain 320 (80 images × 4 lines) and 80 (20 images × 4 lines) spike images, and 4,638 and 1,134 labeled spikelets, respectively. The loss for the validation dataset was calculated, and the precision (AP@0.5IOU) was selected after each iteration to monitor the training process of the Faster RCNN model. The validation loss and AP@0.5IOU are summarized in **Figure 5**.

The AP@0.5IOU of the obtained model was not high due to the inaccuracy and imperfection of the initial labeling datasets; therefore, the training was stopped early when the AP@0.5IOU was essentially unchanged, which was about 0.7479 at iteration 5,393. The final loss of the model for the validation dataset was ∼0.9886.

The training and validation datasets were processed using the obtained model to detect their spikelets. The labeled spikelet numbers for the wheat lines using the model are also summarized in **Table 2**. After the DCNN model labeling, the number of labeled spikelets for the training and validation datasets were 3,973 and 955, respectively. Although the number of labeled spikelets declined, the labeling qualities of spikelets were improved. The boundaries of spikelets were flagged more accurately. In the **Supplementary Material**, the labeled spikelets for each wheat line are detailed.

According to the detection results, we found that the robustness of the obtained model for the spikelets at the bottoms of wheat spikes (as the lowest spikelet in **Figure 4D**) is poor. In practice, these spikelets are sterile and not taken into account. A manual correction was conducted to remove the mislabeled samples. Besides, the model cannot detect all spikelets because some spikelets in the initial datasets were not labeled. Bounding boxes were manually added to handle this situation. The final labeled spikelet samples after manual correction are summarized in **Table 2**, and the training and validation datasets contain 5,024 and 1,204 labeled spikelets, respectively.

Spikelet Detection and Counting Results

Model Training for Spikelet Detection

The Faster RCNN model was retrained using the corrected training and validation datasets. The recorded losses and AP@0.5IOU are summarized in **Figure 6**. Results showed that the training and validation losses of the model decreased slowly with increasing iterations. AP@0.5IOU reached a high value after a few hundred iterations and basically kept stable. Combing AP@0.5IOU with the training and validation losses as the indicators, the model saved at iteration 8,026 was selected as the reference model, which was used to detect and count the spikelets in spike images. At this iteration, the training loss, validation loss, and AP@0.5IOU were about 0.5389, 0.9108, and 0.9582, respectively.

Spikelet Detection and Counting Results

To evaluate the performance of the retrained Faster RCNN model, the testing dataset consisting of 170 spike images for each wheat line that were not selected to generate the training and validation datasets was used to assess the detection and counting methods. The spikelets for each wheat line were detected (**Figure 7**) and counted using the retrained model. The results were summarized and compared with manual counting, as shown in **Figures 8**, **9**. The average processing time of spike images required for spikelet counting was ∼0.4 s. Most absolute errors for all wheat lines were not more than 1. The *RMSE*, *rRMSE*, and R^2 between the automatically and manually counted results for Shiyou 20, Shannong 25, Liangxing 99, and Shenmai 818 varieties were 0.62, 0.58, 0.54, and 0.77; 3.96, 3.73, 3.34, and 4.94%; and 0.73, 0.78, 0.84, and 0.67, respectively. The detection results for all wheat lines are provided in the **Supplementary Material**.

DISCUSSION

Image Annotation

A method based on a watershed algorithm was proposed to assist in the labeling of spikelet samples. Cb component of the spike in the color image was applied to obtain the gray image of the spike. The color component makes full use of the color characteristics of wheat glumes. On this basis, the watershed algorithm can roughly segment the spike into many regions that contain spikelets. To a certain extent, the application of MER reduces the segmentation errors produced by the watershed algorithm. In addition, the effect of wheat awns on spikelet labeling is small. The areas and ratio of length to width were set for MER, which can eliminate the regions that contain wheat awns. Although the proposed method successfully labeled many spikelets, it can achieve better performance on the wheat line whose color distribution is relatively uniform, such as Liangxing 99.

Manual corrections were conducted twice. For the first time, labeled spike images with higher annotation accuracy were manually selected. For the second time, more work was done to supplement unlabeled spikelets, modify some bounding boxes that were with small errors, and remove the labels for the sterile spikelets at the bottom of wheat spikes. Manual corrections assist us to improve the accuracy of annotation.

In the ACID dataset, Pound et al. (2017) annotated each spikelet by placing a dot in its center. In the present study, the spikelets were labeled using bounding boxes. The bounding boxes cannot only help in the detection and counting of the spikelets, but also facilitate the calculation of the length, width, and area of the spikelet in the present study.

Labeling the datasets by integrating machine learning algorithms and deep learning techniques can effectively reduce the labor cost and obtain high-quality datasets. The annotated datasets can be used for DCNN model training.

Spikelet Detection and Counting

In this study, a Faster RCNN model was applied to detect and count the spikelets of the wheat spike. In the image detection field, image resolution has an impact on the detection

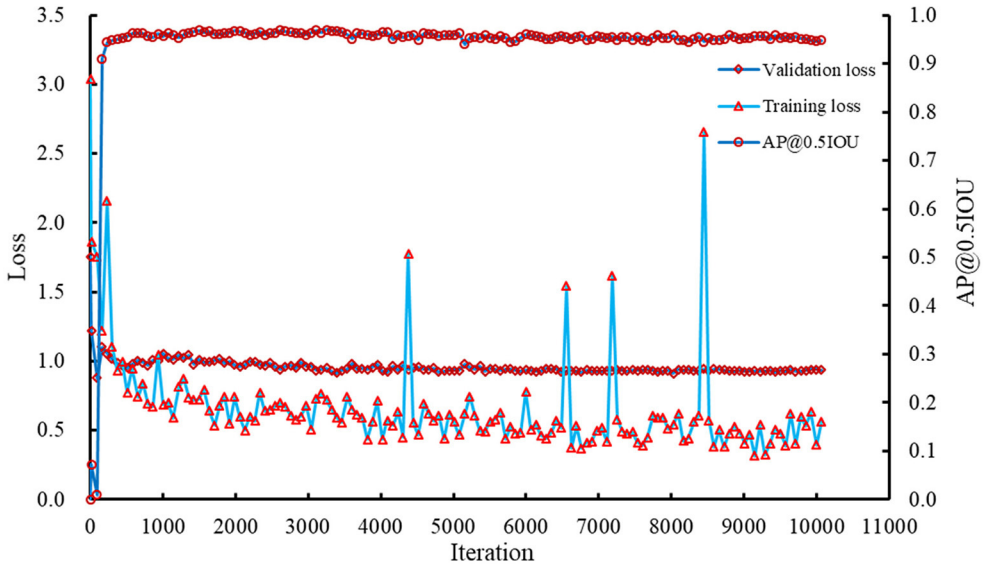


FIGURE 6 | Losses and AP@0.5IOU of the Faster RCNN model during the training process.

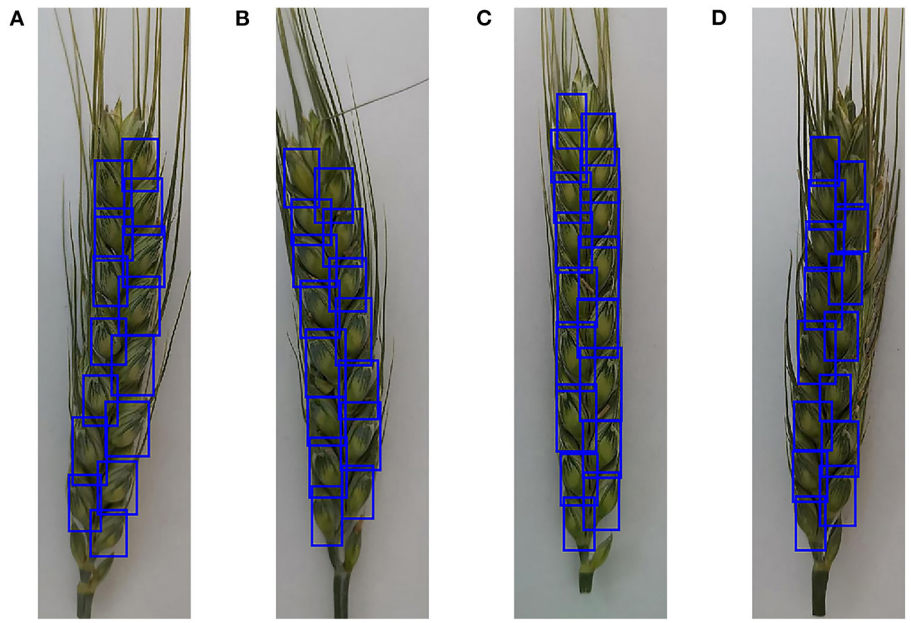


FIGURE 7 | Examples to show the detection results of spikelets for different wheat lines. **(A)** Detection results of Shiyou 20, **(B)** detection results of Shannong 25, **(C)** detection results of Liangxing 99, and **(D)** detection results of Shenmai 818.

performance of the model. During the training process, high image resolution requires expensive processes and higher hardware specifications. To detect the spikes in color images with high resolution, Madec et al. (2019) split the spike images into multiple sub-images and kept a 50% overlap between the sub-images to develop datasets for model training. The sub-images were processed, and spikes were detected to generate bounding boxes for spikes. The overlapped ratios of bounding boxes for neighbored sub-images were judged to remove some repeated detections. An object can be easily detected more than once using this strategy, and the detection results are highly dependent on the sub-images and overlapped ratio setting. Therefore, we prefer to reduce the image resolution to get satisfactory images. Although the ratio of length to width for the image is not 1:1 and is nearly 4:1, which is limited by the spike shape, the performance of our model for spikelet detection and counting shows that

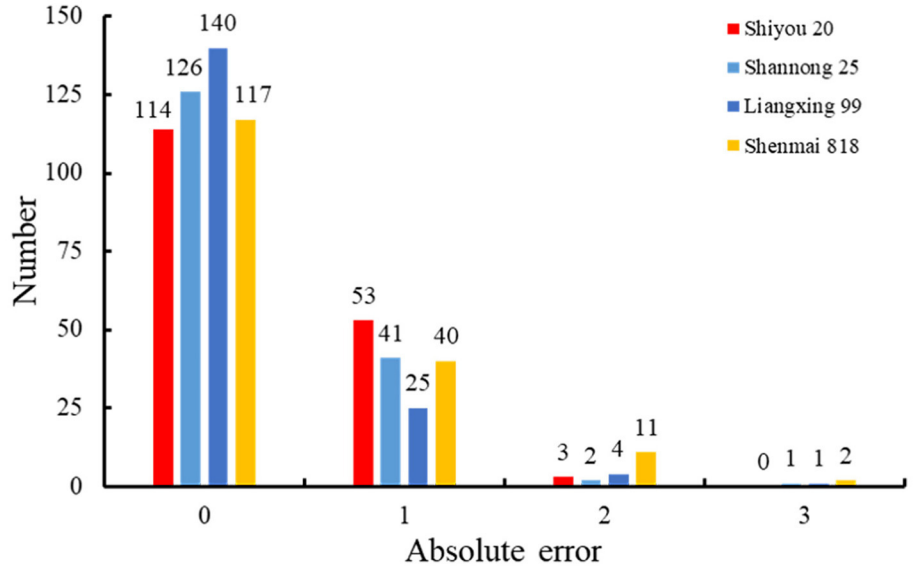


FIGURE 8 | Distribution of the spikelet counting errors for all wheat lines.

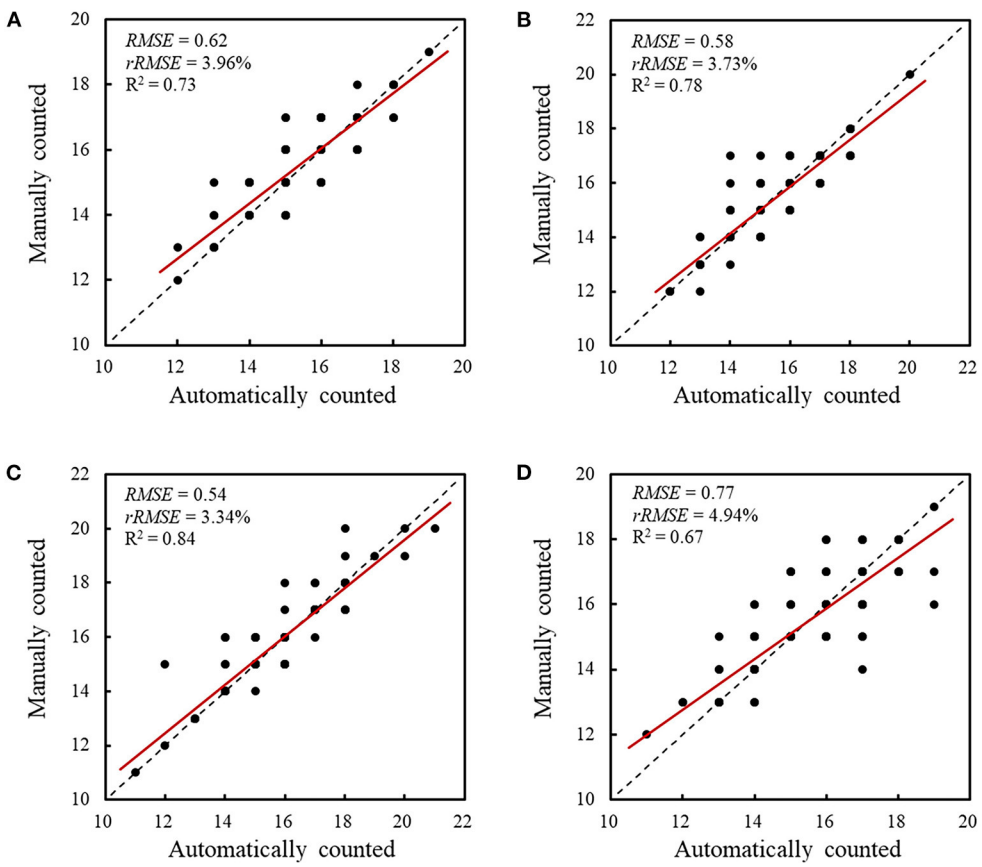


FIGURE 9 | Comparison between the automatically and manually counted spikelet results. **(A)** Comparison and correlation for Shiyou 20, **(B)** comparison and correlation for Shannong 25, **(C)** comparison and correlation for Liangxing 99, and **(D)** comparison and correlation for Shenmai 818. The red lines indicate the least squares linear regression lines. The dashed lines are 1:1 lines.

this strategy is appropriate for processing the color image that contains a single spike.

Pound et al. (2017) conducted similar work to develop models for spikelet counting using manually labeled samples. Their spikelet counting errors ranged from 0.06 to 3.81%. In our study, the spikelet counting and absolute errors were found to be 0.23 and 2.00%, respectively. There were no significant differences between our studies. Therefore, the proposed labeling method and developed model are practical for spikelet counting.

The spikelet detection and counting accuracy of our system is negatively affected by some factors. First and foremost, the testing results demonstrate that the sterile spikelets at the bottom of wheat spikes were mistakenly taken into account. Although the spikelet datasets were manually corrected and the sterile spikelet labels were removed, the sterile spikelets are not uniform for different wheat lines. First, some wheat lines hardly produce sterile spikelets. Second, even for the same wheat line, the sterile spikelets are missing in different spike samples. Third, there are great differences in the locations of the sterile spikelets. Many of them lay next to the spikelets that need to be counted, while others are independent. The performance of the final DCNN model for the sterile spikelet discrimination has been significantly improved compared to the model trained using the initial image annotation. Further work can be conducted to improve the sterile spikelet detection by combing its location and shapes (e.g., area, width, and the ratio of length to width). Another potential problem is that many spikelets are too big to be counted twice. Some grains inside the spikelets are plump, which makes the lateral florets prominent, due to which the lateral florets were incorrectly identified as spikelets. Finally, the effects of wheat awns on spikelet detection should be considered. Some spikelets were covered by awns in some spikes, particularly the upper spikelets, which were hard to be detected. In the following work, the covered spikelets may be estimated according to the symmetrical characteristics of spikelets.

CONCLUSION

In this study, novel methods using color component selection and image processing techniques, combined with deep learning, were proposed to detect and count wheat spikelets in color images. Cb component and a watershed algorithm were implemented to process the color images of the spike and automatically label the spikelets. A DCNN model that was trained using the initially labeled datasets can further enlarge and optimize the datasets. The proposed labeling method can improve the efficiency and accuracy of dataset annotation. Then, a Faster RCNN model, retrained through the transfer learning technique and the obtained datasets, was capable of detecting and counting the spikelets in a spike image. For four wheat lines, *RMSE*, *rRMSE*, and R^2 for the automatic and manual countings of spikelets were 0.62, 0.58, 0.54, and 0.77; 3.96, 3.73, 3.34, and 4.94%; and 0.73, 0.78, 0.84, and 0.67, respectively. These results demonstrated that the proposed methods can effectively detect and count spikelets, which will help breeders collect sufficient data to analyze the developmental characteristics of wheat spikes.

In future work, color images of spikes of other wheat lines will be collected to further test the applicability of the proposed methods. The detected spikelets contained several sterile spikelets at times. Hence, a model that can recognize sterile spikelets needs to be developed. The results of this study can also be combined with other wheat spike properties (e.g., GNS) to evaluate more wheat traits, such as spikelet fertilities. In addition, an app for smartphones can be developed to acquire the number of spikelet samples for field observation.

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

RQ: methodology, software, and writing of the original draft. YH: supervision and fund acquisition. MZ: writing, reviewing, and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Key Projects of International Scientific and Technological Innovation Cooperation among Governments under National Key R&D Plan (2019YFE0103800).

ACKNOWLEDGMENTS

The authors are grateful to Youbing Wang, University of Minnesota, for his instructions in deep learning.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.872555/full#supplementary-material

REFERENCES

Alkhudaydi, T., Zhou, J., and De La lglesia, B. (2019). "SpikeletFCN: counting spikelets from infield wheat crop images using fully convolutional networks," in *International Conference Artificial Intelligence and Soft Computing* (Zakopane), 3–13.

Ayalew, T. W., Ubbens, J. R., and Stavness, I. (2020). "Unsupervised domain adaptation for plant organ counting," in *Proceedings of European Conference on Computer Vision (ECCV)* (Glasgow), 330–346.

Chen, F., Pound, M. P., and French, A. P. (2021). "Learning to localise and count with incomplete dot-annotations," in *Proceedings of the IEEE International Conference on Computer Vision* (Montreal, QC), 1612–1620.

Duan, L., Yang, W., Bi, K., Chen, S., Luo, Q., and Liu, Q. (2011). Fast discrimination and counting of filled/unfilled rice spikelets based on bi-modal imaging. *Comput. Electron. Agric.* 75, 196–203. doi: 10.1016/j.compag.2010.11.004

García, G. A., Serrago, R. A., González, F. G., Slafer, G. A., Reynolds, M. P., and Miralles, D. J. (2014). Wheat grain number: identification of favourable physiological traits in an elite doubled-haploid population. *F. Crop. Res.* 168, 126–134. doi: 10.1016/j.fcr.2014.07.018

Genaev, M. A., Komyshev, E. G., Smirnov, N. V., Kruchinina, Y. V., Goncharov, N. P., and Afonnikov, D. A. (2019). Morphometry of the wheat spike by analyzing 2D images. *Agronomy* 9, 390. doi: 10.3390/agronomy9070390

Giuffrida, M. V., Dobrescu, A., Doerner, P., and Tsaftaris, S. A. (2019). "Leaf counting without annotations using adversarial unsupervised domain adaptation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2590–2599.

Guo, Z., Zhao, Y., Röder, M. S., Reif, J. C., Ganal, M. W., Chen, D., et al. (2018). Manipulation and prediction of spike morphology traits for the improvement of grain yield in wheat. *Sci. Rep.* 8, 14435. doi: 10.1038/s41598-018-31977-3

Hu, G., Qian, L., Liang, D., and Wan, M. (2019). Self-adversarial training and attention for multi-task wheat phenotyping. *Appll. Eng. Agric.* 35, 1009–1014. doi: 10.13031/aea.13406

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., et al. (2017). "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings 30th IEEE Conference Computer Vision and Pattern Recognition*, 3296–3305.

Hughes, N., 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

Kaya, E., and Saritas, I. (2019). Towards a real-time sorting system: identification of vitreous durum wheat kernels using ANN based on their morphological, colour, wavelet and gaborlet features. *Comput. Electron. Agric.* 166, 105016. doi: 10.1016/j.compag.2019.105016

Khoroshevsky, F., Khoroshevsky, S., and Bar-Hillel, A. (2021). Parts-per-object count in agricultural images: solving phenotyping problems *via* a single deep neural network. *Remote Sens.* 13, 2496. doi: 10.3390/rs13132496

Li, Q., Cai, J., Berger, B., Okamoto, M., and Miklavcic, S. J. (2017). Detecting spikes of wheat plants using neural networks with Laws texture energy. *Plant Methods* 13, 83. doi: 10.1186/s13007-017-0231-1

Liu, T., Chen, W., Wang, Y., Wu, W., Sun, C., Ding, J., et al. (2017). Rice and wheat grain counting method and software development based on Android system. *Comput. Electron. Agric.* 141, 302–309. doi: 10.1016/j.compag.2017.08.011

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.* 264, 225–234. doi: 10.1016/j.agrformet.2018.10.013

Misra, T., Arora, A., Marwaha, S., Chinnusamy, V., Rao, A. R., Jain, R., et al. (2020). SpikeSegNet-a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* 16, 40. doi: 10.1186/s13007-020-00582-9

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*, 2055–2063.

Qiu, R., Wei, S., Zhang, M., Li, H., Sun, H., Liu, G., et al. (2018). Sensors for measuring plant phenotyping: a review. *Int. J. Agric. Biol. Eng.* 11, 1–17. doi: 10.25165/j.ijabe.20181102.2696

Qiu, R., Yang, C., Moghimi, A., Zhang, M., Steffenson, B. J., and Hirsch, C. D. (2019). Detection of Fusarium head blight in wheat using a deep neural network and color imaging. *Remote Sens.* 11, 2658. doi: 10.3390/rs11222658

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

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., and Hawkesford, M. J. (2019). DeepCount: in-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks. *Front. Plant Sci.* 10, 1176. doi: 10.3389/fpls.2019.01176

Vahamidis, P., Karamanos, A. J., and Economou, G. (2019). Grain number determination in durum wheat as affected by drought stress: an analysis at spike and spikelet level. *Ann. Appl. Biol.* 174, 190–208. doi: 10.1111/aab.12487

Xiong, B., Wang, B., Xiong, S., Lin, C., and Yuan, X. (2019a). 3D morphological processing for wheat spike phenotypes using computed tomography images. *Remote Sens.* 11, 1110. doi: 10.3390/rs11091110

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., and Shen, C. (2019b). TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* 15, 150. doi: 10.1186/s13007-019-0537-2

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., et al. (2020). Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant Methods* 16, 106. doi: 10.1186/s13007-020-00648-8

Yu, L., Shi, J., Huang, C., Duan, L., Wu, D., Fu, D., et al. (2021). An integrated rice panicle phenotyping method based on X-ray and RGB scanning and deep learning. *Crop J.* 9, 42–56. doi: 10.1016/j.cj.2020.06.009

Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J., and Yang, G. (2018). Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. *Front. Plant Sci.* 9, 1024. doi: 10.3389/fpls.2018.01024

Zhou, H., Riche, A. B., Hawkesford, M. J., Whalley, W. R., Atkinson, B. S., Sturrock, C. J., et al. (2021). Determination of wheat spike and spikelet architecture and grain traits using X-ray Computed Tomography imaging. *Plant Methods* 17, 26. doi: 10.1186/s13007-021-00726-5

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 June 2022
doi: 10.3389/fpls.2022.872107

Deep Learning Based Automatic Grape Downy Mildew Detection

Zhao Zhang[1,2,3,4], *Yongliang Qiao*[5]*, *Yangyang Guo*[1,3,4] *and Dongjian He*[1,3,4]*

[1] College of Mechanical and Electronic Engineering, Northwest A&F University, Xianyang, China, [2] College of Electronic and Electrical Engineering, Baoji University of Arts and Sciences, Baoji, China, [3] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Northwest A&F University, Xianyang, China, [4] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Northwest A&F University, Xianyang, China, [5] Faculty of Engineering, Australian Centre for Field Robotics (ACFR), The University of Sydney, Sydney, NSW, Australia

Grape downy mildew (GDM) disease is a common plant leaf disease, and it causes serious damage to grape production, reducing yield and fruit quality. Traditional manual disease detection relies on farm experts and is often time-consuming. Computer vision technologies and artificial intelligence could provide automatic disease detection for real-time controlling the spread of disease on the grapevine in precision viticulture. To achieve the best trade-off between GDM detection accuracy and speed under natural environments, a deep learning based approach named YOLOv5-CA is proposed in this study. Here coordinate attention (CA) mechanism is integrated into YOLOv5, which highlights the downy mildew disease-related visual features to enhance the detection performance. A challenging GDM dataset was acquired in a vineyard under a nature scene (consisting of different illuminations, shadows, and backgrounds) to test the proposed approach. Experimental results show that the proposed YOLOv5-CA achieved a detection precision of 85.59%, a recall of 83.70%, and a mAP@0.5 of 89.55%, which is superior to the popular methods, including Faster R-CNN, YOLOv3, and YOLOv5. Furthermore, our proposed approach with inference occurring at 58.82 frames per second, could be deployed for the real-time disease control requirement. In addition, the proposed YOLOv5-CA based approach could effectively capture leaf disease related visual features resulting in higher GDE detection accuracy. Overall, this study provides a favorable deep learning based approach for the rapid and accurate diagnosis of grape leaf diseases in the field of automatic disease detection.

Keywords: grape downy mildew, disease detection, deep learning, attention mechanism, data augmentation, digital agriculture

OPEN ACCESS

Edited by:
Yiannis Ampatzidis,
University of Florida, United States

Reviewed by:
Brun Francois,
Association de Coordination Technique Agricole, France
Harald Scherm,
University of Georgia, United States

***Correspondence:**
Yongliang Qiao
y.qiao@acfr.usyd.edu.au
Dongjian He
hdj168@nwsuaf.edu.cn

Specialty section:
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

Received: *09 February 2022*
Accepted: *27 April 2022*
Published: *09 June 2022*

Citation:
Zhang Z, Qiao Y, Guo Y and He D (2022) Deep Learning Based Automatic Grape Downy Mildew Detection. Front. Plant Sci. 13:872107. doi: 10.3389/fpls.2022.872107

1. INTRODUCTION

Grape as an important fruit crop makes a large economic income contribution in many countries (Liu et al., 2020; Zhou et al., 2021). As the grape grows in a natural condition, diseases will often appear on the leaves due to the complex weather condition and changing surrounding environments. Grape downy mildew (GDM) is one of the serious diseases caused by the oomycete pathogen Plamopara viticola, which seriously affects the growth of the grapes, causes a decrease in quality and yield, and results in huge economic losses in the grape industry (Chen et al., 2020; Ji et al., 2020). Downy mildew often happened in wet and rainy areas in spring and summer, it is initiated at the stomata on the underside of the leaf, and then on the whole leaf

(Chen et al., 2020). Monitoring grape leaf health and detecting pathogen are essential to reduce disease spread and facilitate effective management practices. Grape leaf diseases are currently controlled by repetitive fungicide treatments throughout the season. Reducing the treatment costs is a major challenge from both environmental and economic views. Timely detection and treatment at the initial stage of downy mildew infection (Adeel et al., 2019) is a good solution to control and cut down the spread of downy mildew in a large area. Therefore, if an automatic detection method can be achieved when the spots appear, the leaf disease control plan can be made to control the diseases, guarantee the grape plant health, and improve the quality and yield of the grapes. Vision based detection approaches have been developed to detect plant diseases, which is performed by extracting visual features (e.g., texture, shape, and color of leaf lesions) of leaf images and using models (e.g., support vector machine, linear regression) to recognize and detect the diseases (Tang et al., 2020; Hernández et al., 2021). Zhu et al. (2020) identified grape diseases using image analysis and BP neural networks. Chen et al. (2020) developed and compared several generalized linear models to predict the probability of high incidence and severity in the Bordeaux vineyard region. Abdelghafour et al. (2020) detected downy mildew symptoms using proximal color imaging and achieved 83% pixel-wise precision. However, the traditional image processing technology needs to manually extract the leaf disease characteristics, which is often time-consuming, and easy to miss the best disease prevention time. In addition, under the nature scene (e.g., different illumination, symptoms, camera viewpoints), classical algorithms or models lack robustness and cannot achieve stable detect performance.

Many scholars have proposed approaches for earlier plant disease detection and monitoring of the disease symptoms (Mutka and Bart, 2015). At the earlier stage, the human-crafted features such as texture, color, or shape characteristics are extracted from RGB or hyper-spectral plant leaf images for identifying the plant diseases (Mahlein, 2016). For example, Atanassova et al. (2019) proposed spectral data based classification models to predict the infection in plants, which achieved over 78% accuracy. Waghmare et al. (2016) proposed an automatic grape diseases detection system using the extracted color Local Binary Pattern (LBP) features. Mohammadpoor et al. (2020) proposed a support vector machine for grape fanleaf virus detection and achieved 98.6% average accuracy. However, this kind of method mostly depends on selected features and their extraction is easy to be influenced by the camera viewpoints, shadows, and lighting.

In recent years, deep learning methods such as convolutional neural networks (CNN) have been widely implemented in leaf disease detection, scene perception, and smart agriculture. Variates CNN based detection methods have been proposed for leaf disease recognition and monitoring (Liu et al., 2020). Ferentinos (2018) proposed convolutional neural network architectures to identify healthy or diseased plants. Arsenovic et al. (2019) developed a two-stage architecture of neural networks to classify plant disease and achieved an accuracy of 93.67%. Zhang et al. (2019a) proposed an AlexNet based cucumber disease identification approach, achieving 94.65% recognition accuracy. Ji et al. (2020) proposed CNN based approach to classify common grape leaf diseases and obtained average classification accuracy of 98.57%. Liu et al. (2020) proposed Inception convolutional neural network (DICNN) for identifying grape leaf diseases and realized an overall accuracy of 97.22% on single-leaf datasets. Thet et al. (2020) used an improved VGG16 model that achieved 98.4% classification accuracy for five different leaf diseases. Tang et al. (2020) classified grape disease types using lightweight convolution neural networks and channel-wise attention, which achieved 99.14% accuracy. Liu and Wang (2020) improved the YOLOv3 model to directly generate the bounding box coordinates for tomato diseases and pests detection, which achieved a detection accuracy of 92.39%. According to these studies, CNNs can learn advanced robust features of leaf diseases directly from original images, outperforming the traditional feature extraction approaches. Yu and Son (2020) proposed a leaf spot attention mechanism to increase apple leaf disease discriminative power and enhance the identification performance. Hernández and Lopez (2020) developed Bayesian deep learning techniques and an uncertainty probabilistic programming approach for plant disease detection.

With the continuous development of smart sensors, big data, and cloud computing, many automatic approaches have been proposed to identify and detect plant leaf diseases (Vishnoi et al., 2021). The rapid development of artificial intelligence and the Internet of Things (IoT) has significantly facilitated automatic disease detection (Zhang et al., 2020). Using deep learning models and noninvasive sensors to identify plant diseases has drawn more attention in the field of precision agriculture and plant phenotyping (Nagaraju and Chawla, 2020; Singh et al., 2020). Hernández et al. (2021) investigated hyperspectral sensing technologies and artificial intelligence applications for assessing downy mildew in grapevine under laboratory conditions. Gutiérrez et al. (2021) differentiated downy mildew and spider mite in grapevine under field conditions using the CNN model. Liu et al. (2021) proposed Hierarchical Multi-Scale Attention Semantic Segmentation (HMASS) to identify GDM infected regions, and the calculated infection severity percentage was highly correlated (R = 0.96) with the human field assessment.

Choi and Hsiao (2021) classified Cassava leaf diseases using the Residual Network. Zhang et al. (2021) developed a multi-feature fusion Faster R-CNN model and achieved 83.34% detection accuracy for soybean leaf disease. Dinata et al. (2021) proposed CNN based approach for 6 types of strawberry disease classification and achieved 63.7% accuracy. Abbas et al. (2021) detected tomato plant disease using transfer learning and C-GAN synthetic images, which achieved 99.51% accuracy. Cristin et al. (2020) proposed a deep neural network based Rider-Cuckoo Search Algorithm and achieved 87.7% plant disease detection accuracy. Roy and Bhaduri (2021) proposed deep learning-based multi-class plant disease and achieved 91.2% mean average precision. However, most of these methods are only tested in experimental situations, which need to be verified on the complex background situation.

Despite deep learning based approaches demonstrating its facilitate in GDM detection, the detection accuracy and speed restricted its application in autonomous viticulture management. Plant leaf disease detection in the real vineyard is facing many challenges, such as the small difference between the lesion area and the background, different scales of the spots, variation of symptoms, and camera viewpoints (Liu and Wang, 2021). Also, light changing in a real complex natural environment further increased the difficulty to achieve high detection accuracy. Therefore, real-time and accurate detection of grape downy mildew is of great significance for the scientific management and control of grape diseases in precision vineyard farming.

Recently, the attention mechanisms such as Squeeze-and-Excitation Networks (SE) (Hu et al., 2018), Convolutional block attention module (CBAM) (Woo et al., 2018), and CA (Hou et al., 2021) have been widely used to enhance the deep learning model performances. SE simply squeezes each 2D feature map to efficiently build interdependencies among channels (Hu et al., 2018). CBAM introduces spatial information encoding via convolutions with large-size kernels and gathers channel-wise and spatial-wise attention sequentially. The recently proposed CA adopts different spatial attention mechanisms and designs advanced attention blocks. Zhang et al. (2019b) applied an attention mechanism to object detection networks, enhancing the impact of significant features and weakening background interference. Experimental results show that the proposed approach achieved an object detection accuracy of 75.9% on PASCAL VOC 2007, which is 6% higher than Faster R-CNN. Liu et al. (2019) presented a deep neural network architecture based on information transmission and attention mechanisms. Zhao et al. (2021) diagnosed tomato leaf disease using an attention module improved network, which achieved 96.81% average identification accuracy on the tomato leaf diseases dataset. Ravi et al. (2021) integrated the attention module into the EfficientNet model to locate and identify the tiny infected regions in the Cassava leaf. The proposed method achieved better performance than non-attention-based CNN pre-trained models. Wang et al. (2021b) proposed a Fine-Grained grape leaf disease recognition method using a lightweight attention network, which can efficiently diagnose orchard grape leaf diseases with low computing cost. The above studies have demonstrated that attention mechanisms could enhance feature extraction ability for leaf disease detection and identification.

In this study, to improve GDM detection accuracy in the natural grape farm environment, we proposed YOLOv5-CA based GDM detection approach by combing YOLOv5 and CA mechanism. Different scales of image features were extracted through CNN layers of YOLOv5, and these features were weighted by CA for GDM detection. By using CA, the features' effectiveness for GDM detection is highlighted and those less effective features are suppressed. The proposed YOLOv5-CA based GDM detection is tested on our acquired grape leaf image dataset.

The remaining part of the article is organized as follows. Section 2 illustrates the used datasets, the proposed approaches, and evaluation indicators. Experimental results are presented in Section 3. Discussions of the performance are presented in Section 4. Finally, conclusions and future areas for research are given in Section 5.

2. MATERIAL AND METHODOLOGY

2.1. Plant Material and Image Acquisition

Grape leaf image data were acquired in a commercial vineyard located in the college of Enology, Northwest A&F University, north of China (Yangling, Shaanxi Province). The vineyard manifested downy mildew (Plasmopara viticola) in many plants. Images were taken manually for several days (each day is from 8:00 a.m. to 16:00 p.m.) in early August on a partly cloudy day (**Figure 1**). The used camera is a Canon EDS 1200D (a field of view of approximately 504 mm horizontally and 360 mm vertically), and the external conditions for shooting are automatic mode. There is approximately 30 cm between the camera lens and the grape leaves.

A total of 820 leaf samples were collected from different lights, leaf overlapping, and disease severity. The dataset is challenging considering the complex background, occlusions, different disease spot-areas, and shadow influence. **Figure 1** shows images of diseased leaves in a typical complex environment in the dataset. Downy mildew first appears as brown patches. These patches gradually spread and a leaf that is severely affected may have a reduced yield with a shorter lifetime and fruits with a small size.

To validate the proposed YOLOv5-CA based GDM detection approach, the randomly selected 500 leaf images were used as training datasets, while the remaining 320 images were used as testing data. For experiment testing, the LabelImg annotation tool (Tzutalin, 2015) was used to manually label the leaf disease areas.

2.2. YOLOv5-CA Based GDM Detection

In order to make YOLOv5 more suitable for GDM detection in complex natural scenarios such as complex background, occlusions, different disease spot-areas, and shadow influence, YOLOv5-CA based GDM detection approach is proposed to improve the GDM detection performance for real farming applications. Grape leaves' RGB images were acquired under field conditions from a commercial vineyard. These collected images contain healthy and downy mildew infected leaves. Then detection model YOLOv5-CA was trained to identify the GDM infected regions. As shown in **Figure 2**, the proposed YOLOv5-CA approach extracted features using YOLOv5 and learned key features through CA, enhancing the feature extraction ability and improving the leaf disease detection performance. As YOLOv5 could adjust the width and depth of the backbone network according to application requirements, for GDM detection, moderate model parameters (i.e., width and depth parameters are 0.75 and 0.67, respectively) were used to achieve reasonable detection speed.

YOLOv5-CA network is mainly composed of backbone part, neck network, and head part: 1) The backbone of YOLOv5 is responsible for extracting image features, which includes several different layers types such as Focus, Conv, C3, CA, and Spatial

FIGURE 1 | Commercial vineyard and acquired images under natural light conditions.

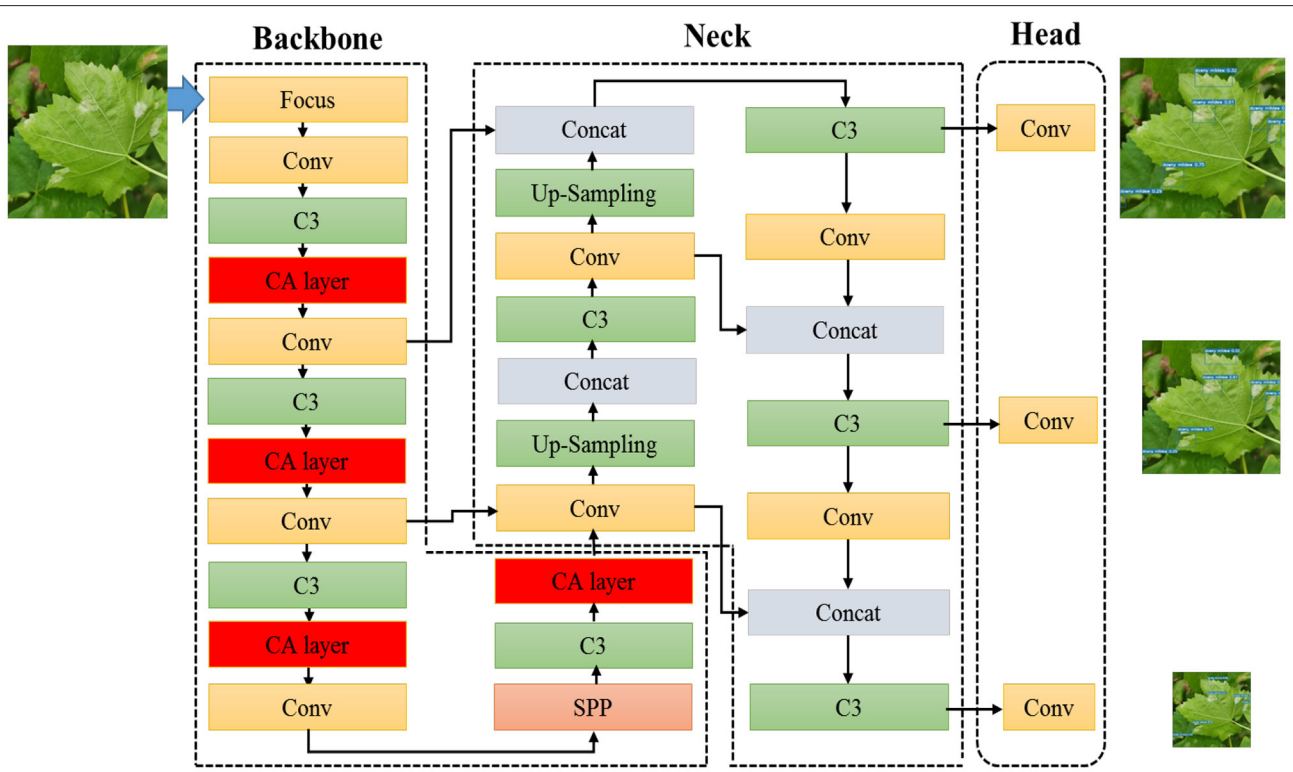


FIGURE 2 | The architecture of the proposed YOLOv5-CA based GDM detection.

Pyramid Pooling (SPP) layer. 2) The neck module generates a feature pyramid based on the PANet (Path Aggregation Network) (Liu et al., 2018). It is a series of feature aggregation layers of mixed and combined image features, enhancing the ability to detect objects with different scales by fusing low-level spatial features and high-level semantic features. 3) The head module generates detection boxes, indicating the category, coordinates, and confidence by applying anchor boxes to multi-scale feature maps from the neck module. The proposed YOLOv5-CA boosts the detection ability of different GDM infection regions through an attention mechanism, which provides a feasible GDM detection and monitoring solution for automatic disease control.

2.2.1. Backbone of YOLOv5-CA

The backbone of the YOLOV5-CA object detector mainly contains Focus, Conv, C3, CA, and Spatial Pyramid Pooling (SPP) layer. The features from deeper layers are more abstract and semantic, while the low-layer features contain spatial information and fine-grained features. For an input image, the Focus module rearranged it through stridden slice operations

in both width and height dimensions, which reduces model calculation time. C3 module contains three convolutions and is used to extract the deep features of the image. The following SPP is used to improve the receptive field of the network by converting any size of the feature map into a fixed-size feature vector. SPP (He et al., 2015) concatenates layer outputs with different kernel sizes (e.g., 13×13, 9×9, 5×5) to boost multi-scale image feature representation ability. All the convolutions utilize Swish activation:

$$Swish(x) = x \times \sigma(x) \tag{1}$$

where σ denotes the sigmoid function.

In our study, we integrated the CA layer into the YOLOv5 backbone, CA layer factorizes channel attention into two 1D feature encoding processes and preserves the precise positional information, which augments the representations of the leaf disease regions.

2.2.2. Neck of YOLOv5-CA

The neck structure used in YOLOv5-CA is a PANet (Liu et al., 2018), which fuses the information of all layers to aggregate features by combing bottom-up pyramid and element-wise max operations. PANet combines convolution features of different layers for images, thus the useful information in each feature layer can be directly propagated to the following subnetwork. By this, PANet can not only realize the abstract description of large objects but also retains the feature details of small objects. In addition, C3 modules are also added at this stage to enhance the feature fusion capability. Through the neck part, the features of infected areas can be extracted to maintain the detection performance.

2.2.3. CA Layer

In terms of GDM detection, because the GDM is randomly distributed in the grape leaf, there is inevitably a mix of overlapping occlusion, and the GDM infection regions account for a relatively small percentage of the images, resulting in missed and mis-detected. In our study, a plug-and-play CA layer was introduced to assist YOLOv5 focused on key disease-related features, and improve the detection accuracy.

The CA layer embeds the location-aware information into the channel attention simultaneously, which increases the spatial range of attention and avoids a lot of computational overhead (Hou et al., 2021). CA layer can be regarded as a computational unit that enhances the representation ability of the learned features. For any intermediate feature $X = [x1, x2, \cdots x_c] \in \mathbb{R}^{C\times H\times W}$, CA could outputs a transformed feature with augmented representations $Y = [y1, y2, \cdots, y_c]$ of the same size to X.

As shown in **Figure 3**, the CA mechanism can be divided into two parts: the coordinate information embedding part (encodes the information of the channels in the horizontal and vertical coordinates) and the coordinate attention generation part (captures the positional information and generates the weight values).

2.2.4. Coordinate Information Embedding

Attention mechanisms have been demonstrated helpful to enhance the overall performance of deep learning models (Chorowski et al., 2015). The attention mechanism can be regarded as a feature weighting scheme, which helps the deep learning model to pay more attention to the task-related information, and suppress or ignore the less-contribution features (Li et al., 2020; Mi et al., 2020). Through this, the attention mechanism strengthens the deep learning model's learning ability and boosts performance (Niu et al., 2021). In recent years, attention mechanisms based on deep learning networks have been applied to a wide variety of computer vision tasks such as image classification, object detection, and image segmentation (Qiao et al., 2019, 2021). Wang et al. (2021a) developed a deep attention module for vegetable and fruit leaf plant disease detection. Kerkech et al. (2020) used a fully convolutional neural network approach to classify Unmanned Aerial Vehicle (UAV) image pixels for detecting mildew disease.

It is known that channel attention could increase the value of the important channel while punishing the non-significant channels, however, channel attention is difficult to preserve positional information (Zhang et al., 2018). To capture precise positional information, the global average pooling was factorized into the average pooling from two directions of each channel. Specifically, given the input X, two spatial extents of pooling kernels (H, 1) and (W, 1) were used to encode each channel along the horizontal coordinate and the vertical dimensions, respectively. The output of the *c-th* channel along height h and width w dimensions can be formulated as:

$$\begin{aligned} z_c^h(h) &= \frac{1}{W}\sum_{0\le i<W} x_c(h,i), \\ z_c^w(w) &= \frac{1}{H}\sum_{0\le j<H} x_c(w,j). \end{aligned} \tag{2}$$

where z^h and z^w are the outputs of the transform at h direction width w, respectively; x_c is the feature map at *c-th* channel; W and H are the width and height dimensions of the feature map separately.

The Equation (2) encodes each channel along with the horizontal and vertical coordinates, preserving the positional information of each channel of feature maps, which facilitates the network to locate the GDM-related visual features precisely.

2.2.5. Coordinate Attention Generation

To further exploit resulting expressive representations, a simple and effective coordinate attention generation was used as the second transformation. Here, the obtained feature maps from the coordinate information embedding stage were concatenated and then sent to a shared 1×1 convolution layer. The relevant process is defined as:

$$f = Relu(F([z^h, z^w])) \tag{3}$$

where [,] indicates concatenate operation, F is 1×1 convolution operation; $f \in \mathbb{R}^{\frac{C}{r}\times(W+H)}$ is the output feature map of the ReLU layer, r is reduction rate.

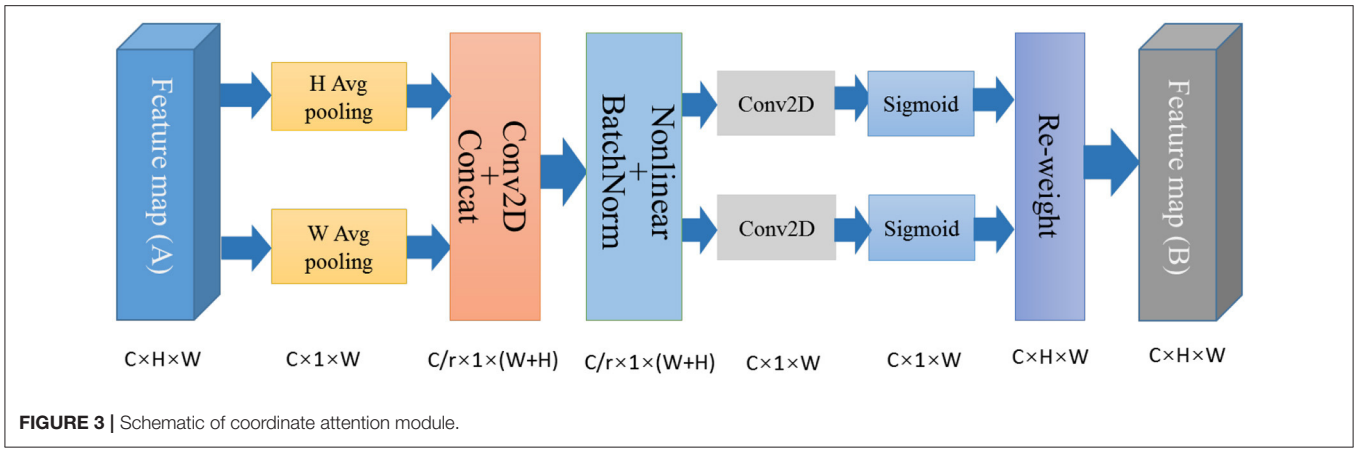


FIGURE 3 | Schematic of coordinate attention module.

Next, the feature map f was decomposed into two separate tensors: $f^h \in \mathbb{R}^{\frac{C}{r}\times H}$ and $f^w \in \mathbb{R}^{\frac{C}{r}\times W}$. Then the following two 1×1 convolution layers for f^h and f^w, respectively, are recovered to the same shape as z^h and z^w. The operation is formulated as:

$$\begin{aligned} g^h &= \sigma(F_h(f^h)), \\ g^w &= \sigma(F_w(f^w)). \end{aligned} \tag{4}$$

where σ is the sigmoid activation function, and F_h and F_w are the convolution manipulation for f^h and f^ω separately.

The obtained feature maps g^h and g^w are then expanded and used as attention weights for the horizontal and vertical coordinates, respectively. This operation can enhance the effective leaf disease related features and reduce the influence of less important information. The reweighing process of the original input feature map can be defined as:

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j). \tag{5}$$

where y_c is the *c-th* channel in the generated feature map y of the attention block.

2.3. YOLOv5-CA Model Training for GDM Detection

2.3.1. Network Training Parameters

In our study, the experimental platform is based on a computer equipped with an NVIDIA RTX 1080Ti GPU, Ryzen 7 3600 CPU@3.6 GHz. The proposed GDM detection approach was implemented using Pytorch.

In addition, to verify the effectiveness of the YOLOv5-CA based GDM detection approach, Faster R-CNN (Ren et al., 2015), YOLOv4 (Bochkovskiy et al., 2020), and YOLOv5 (Tzutalin, 2015) were also used for comparison. Faster R-CNN generates regions of interest (RoIs) candidates and then classifies them into objects (and background) and refines the boundaries of those regions. YOLOv4 and YOLOv5 are the two widely used detection methods from the YOLO series (Redmon et al., 2016).

For network training, the network's input size was set to 416×416×3, the training epoch was set to 1000, batch size was set to 16, and the learning rate was 0.0013. The momentum factor (momentum) was set to 0.937, the initial learning rate was 1×10^{-5} and the decay rate of weight was set to 0.001. The other parameters of each network are their default settings. In the training process, the network predicts the bounding box based on the initial anchor box. The gap between the prediction and ground truth was calculated to update the network in reverse and adjusts the network parameters. After training, the weight file of the detection model obtained was saved.

2.3.2. Network Loss Function

YOLOV5-CA automatically updates the best bounding box for GDM detection during the training process. The default optimization method of the model is the gradient descent method. The loss function L_{loss} used in YOLOv5-CA includes bounding box location loss L_{Ciou}, confidence loss L_{conf} and classification loss L_{cls}:

$$L_{loss} = L_{cls} + L_{conf} + L_{CIoU} \tag{6}$$

Classification loss L_{cls} computes the loss of class probability using Cross Entropy:

$$L_{cls} = \sum_{i=0}^{s^2} \ell_{i,j}^{obj} \sum_{c\in classes} \left[\hat{p}_i(c) log(p_i(c)) + (1-\hat{p}_i(c)) log(1-p_i(c))\right] \tag{7}$$

where ℓ_i^{obj} is used to judge whether there is an object center. $\hat{p}_i(c)$ is the probability of class c; $p_i(c)$ is the probability of predicted box that belongs to class c.

Confidence loss L_{conf} penalizes object confidence error if that predictor is responsible for the ground truth box, which is computed using mean squared error:

$$\begin{aligned} L_{conf} &= \sum_{i=0}^{s^2}\sum_{j=0}^{B} \ell_{i,j}^{obj} \left[\hat{C}_i log(C_i) + (1-\hat{C}_i) log(1-C_i)\right] \\ &+\lambda_{noobj} \sum_{i=0}^{s^2}\sum_{j=0}^{B} \ell_i^{noobj} \left[\hat{C}_i log(C_i) + (1-C_i) log(1-\hat{C}_i)\right] \end{aligned} \tag{8}$$

where λ_{noobj} represents the weight of the classification error, S is the number of grids, and B is the number of prior boxes in each

grid; C_i is the confidence of the predicted box; $\hat{C}_i$ is the confidence of the ground-truth ($\hat{C}_i$ is always 1).

The L_{Ciou} computes the loss related to the predicted bounding box and ground truth, it can be defined as follows:

$$\begin{cases} L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{e^2} + \frac{v^2}{(1-IoU)+v} \\ IoU = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|} \end{cases} \quad (9)$$

where v represents the coincidence degree of the two frame aspect ratios, b and b^{gt} are the center coordinates of the prediction box and the real box respectively; ρ is the Euclidean distance between the two center points, and e represents the diagonal distance of the smallest closed area containing both the prediction and real boxes. IoU means the ratio of the intersection and union of the prediction bounding box and the actual bounding box.

2.3.3. Performance Evaluation

The used performance evaluation indicators for GDM detection include precision, recall, F_1-score, mAP (mean average precision), and FPS (frame per second). Precision shows the ability of the model to accurately identify targets; recall reflects the ability of the model to detect targets; the F_1-score is a harmonic mean of the precision and recall; FPS is the average inference speed. The F_1-score is the reconciled mean of precision and recall, taking into account both the precision and recall of the classification model. Based on tp (the number of hlcorrectly detected downy mildew areas), fp (the number of incorrectly detected downy mildew areas), and fn (the number of disease regions that are incorrectly identified as background), the relevant calculation equations are as follows:

$$Precision = \frac{tp}{tp + fp} \times 100\% \quad (10)$$

$$Recall = \frac{tp}{tp + fn} \times 100\% \quad (11)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (12)$$

From the values of precision and recall, a precision-recall curve can be plotted to observe their distribution. The value of AP is the area under the precision-recall curve, and a larger value means better model performance. mAP@0.5 is the average value of precision under different recall values when the intersection over union (IoU) is 0.5. The calculation of mAP is as follows:

$$mAP = \frac{1}{n} \sum_{k=1}^{N} AP_k \quad (13)$$

where N denotes the number of disease types (N is 1 in our study).

TABLE 1 | Comparison of different GDM methods.

Method	Precision (%)	Recall (%)	F_1 (%)	mAP@0.5 (%)	FPS (Frame/s)
Faster R-CNN	79.97	87.80	83.70	80.65	35.90
YOLOv4	82.69	83.63	83.15	82.65	75.20
YOLOv5	85.35	81.45	83.36	87.41	84.74
YOLOv5-CA	85.59	83.70	84.63	89.55	58.82

3. EXPERIMENTAL RESULTS

3.1. Comparison of Different Object Detection Algorithms

There are varieties of deep learning based detection methods, in order to verify the effectiveness of the proposed method for GDM detection, three popular detection algorithms—Faster R-CNN, YOLOv4, and YOLOv5 were compared. The GDM detection results were presented in **Table 1**.

In **Table 1**, the proposed YOLOv5-CA based approach achieved 85.59% precision, 83.70% recall and 84.63% F_1, and 89.55% mAP, respectively. Compared with the other methods, the proposed YOLOv5-CA GDM detection method is better than that of Faster R-CNN (80.65% mAP), YOLOv4 (82.65 % mAP), and YOLOv5 (87.41% mAP). From these results, it is clear that the CA mechanism of YOLOv5-CA improves the feature representation ability, enhancing the final detection accuracy for identifying the leaf disease areas. Meanwhile, the proposed approach could detect the GDM with a speed of 58.82 frames per second. These results illustrated that the proposed method could achieve high precision with a fast speed to meet real-time requirements, which is favorable for the deployment of the GDM detection model in spraying robots for the plant diseases control in smart vineyard farming.

3.2. Qualitative GDM Detection Comparison

Figure 4 demonstrates the comparison of different methods' qualitative results on our acquired grape leaf dataset. It can be seen that the proposed YOLOv5-CA could detect GDM at different leaf parts (e.g., leaf edge, the leaf central parts). Especially, the proposed YOLOv5-CA method could detect less obvious GDM lesions on the leaves, which outperformed the other methods such as Faster R-CNN, YOLOv4, and YOLOv5. The main reason could be that the CA mechanism strengthens the feature representation ability, which enhances the GDM detection performance.

Additionally, more examples of YOLOv5-CA based GDM detection are presented in **Figure 5**. It can be seen that the GDM infected regions are well detected (blue bounding box) under complex background, especially, YOLOv5-CA could well detect the GDM regions nearby the leaf edge and petioles. It also can be noted that the YOLOv5-CA could detect both large and small GDM regions. The main reason is that the YOLOv5-CA makes the network pay more attention to the GDM-related visual features, reducing the false or mis-detection cases. The

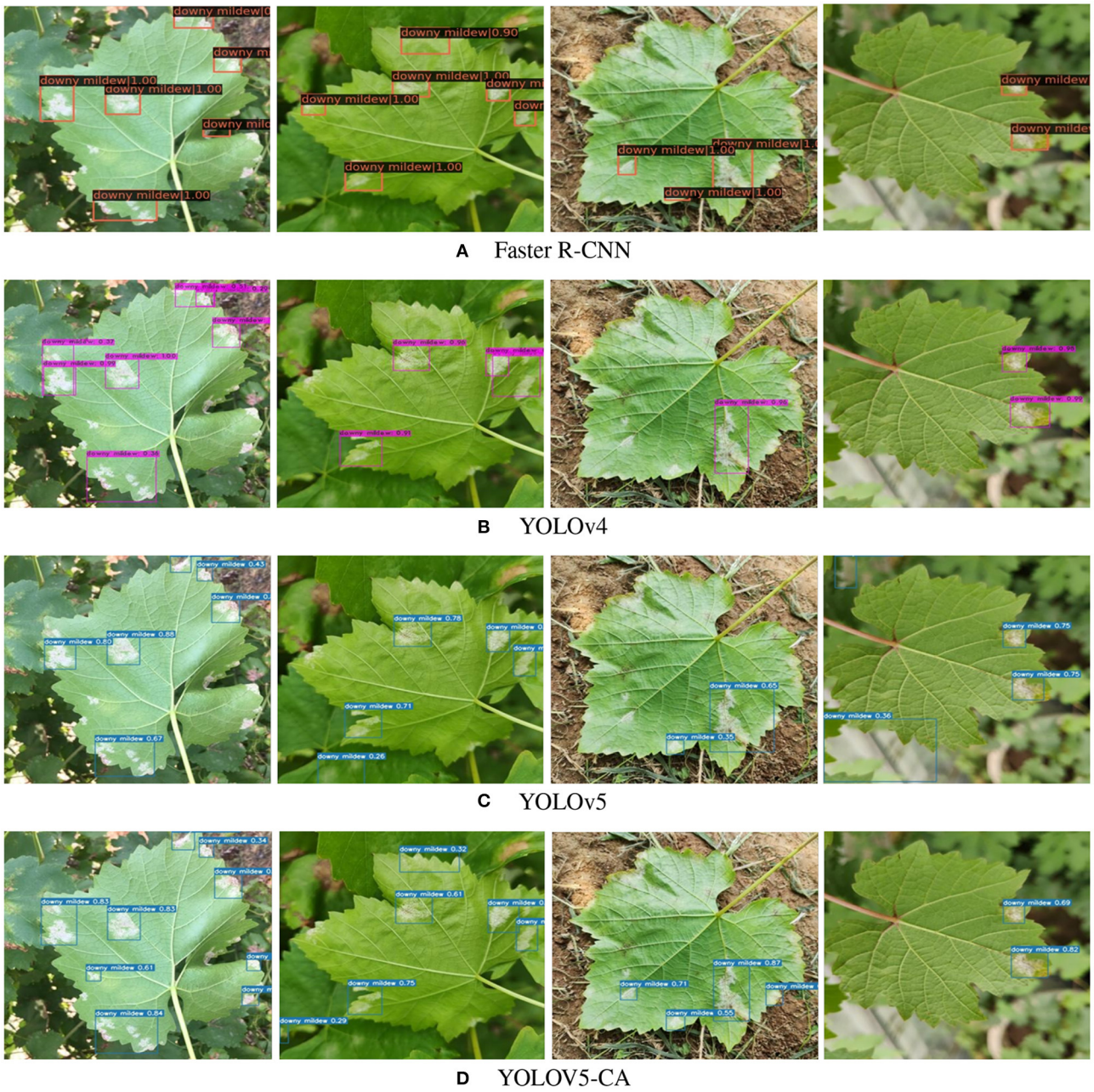


FIGURE 4 | Examples of different GDM detection attention methods. **(A)** Faster R-CNN, **(B)** YOLOv4, **(C)** YOLOv5, and **(D)** YOLOV5-CA.

good detection performance of YOLOv5-CA provides valuable information for automatic disease control.

3.3. Influence of Different Network Input-Sizes on GDM Detection

The network input size is one factor that would influence the GDM detection performance. Here, we also investigate different input-sizes' influence on YOLOv5-CA based GDM detection. In **Table 2**, five typical network input sizes, namely, 112 × 112, 224 × 224, 320 × 320, 416 × 416, and 512 × 512 were compared in terms of GDM detection performance.

According to **Table 2**, the network input with 416×416 size achieved 85.59% precision, 83.70% recall, 84.63% F_1-score, and 89.55% mAP@0.5, which outperformed the performance of input size with 112×112, 224×224, and 512×512. This means the proposed YOLOv5-CA could extract and learn the more useful information from the large input size. However, when the network input-size increases to 512×512, there is not much performance improvement but significantly increased

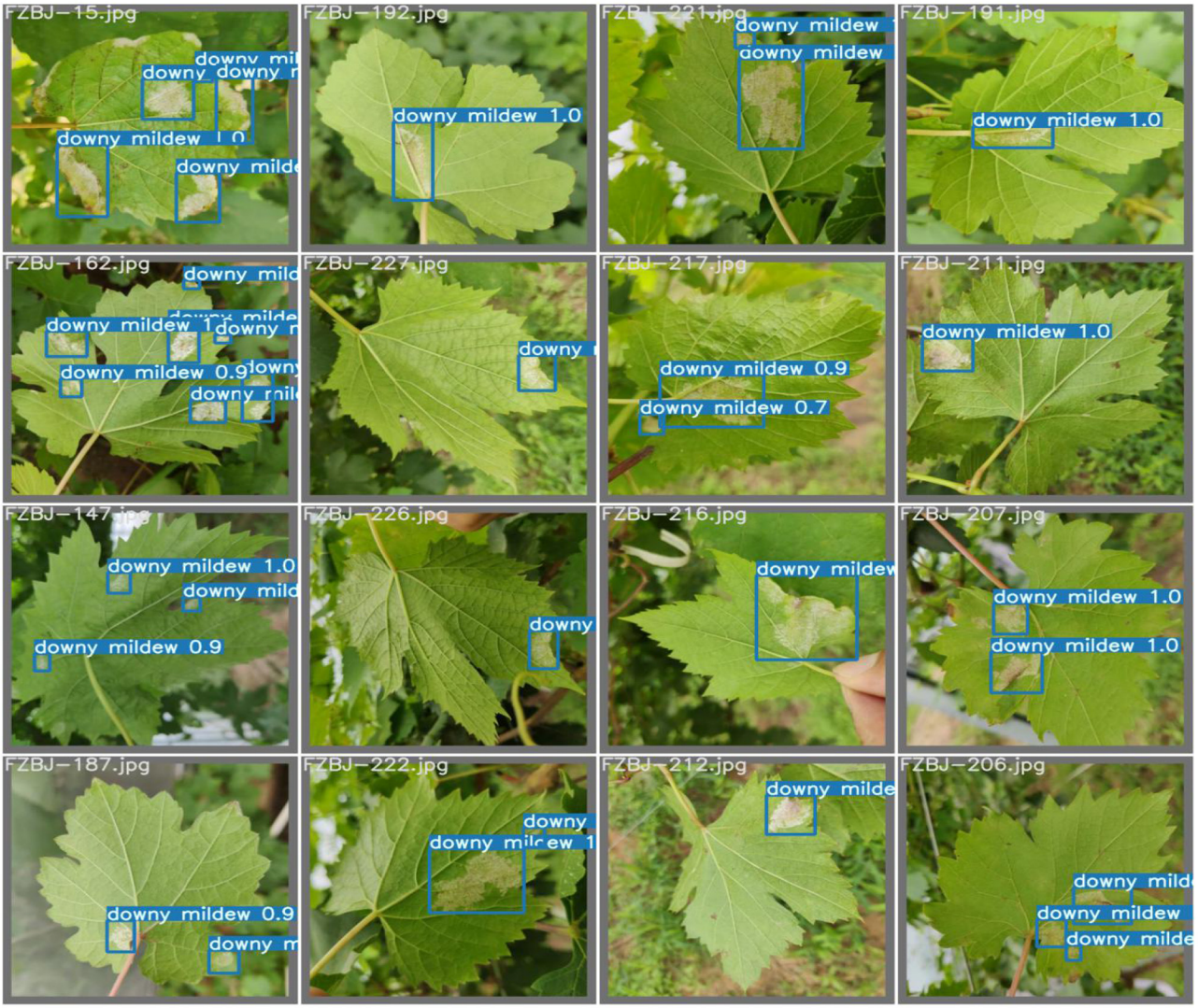


FIGURE 5 | Examples of YOLOv5-CA based GDM detection results.

TABLE 2 | Grape downy mildew Detection performance with different network input sizes.

Network input size	Precision (%)	Recall (%)	F_1 (%)	mAP@0.5	FPS (Frame/s)
112 × 112	80.32	72.76	76.35	76.71	102.04
224 × 224	83.73	79.32	81.47	82.63	92.63
320 × 320	84.75	84.32	84.53	85.25	76.92
416 × 416	85.59	83.70	84.63	89.55	58.82
512 × 512	86.71	82.80	84.71	87.89	45.45

the processing time and calculating memory size, which is not favorable for fast detection and real applications. By balancing the speed and accuracy, the input size of 416×416 was selected in our work for real-time GDM detection.

3.4. Data Augmentation for YOLOv5-CA Detection

Offline data augmentation could increase the dataset diversity, explore the network hyperparameters, and finally enhance the accuracy and robustness of the trained model (Zoph et al., 2020; Su et al., 2021). To further improve the GDM detection performance, in our study, bounding box based data augmentation was used. The augmentation technique was only applied to disease areas within the manually labeled bounding boxes. The transformations for data augmentation implemented include: flipping horizontally and vertically, randomly cropping between 0 and 20% of the bounding box, random rotation, random shear of between −15° to +15° horizontally and vertically, random brightness adjustment (between 0 and +10%), and Gaussian blur (between 0 and 5 pixels). The original 500 training images were expanded to 2000 images, and then they

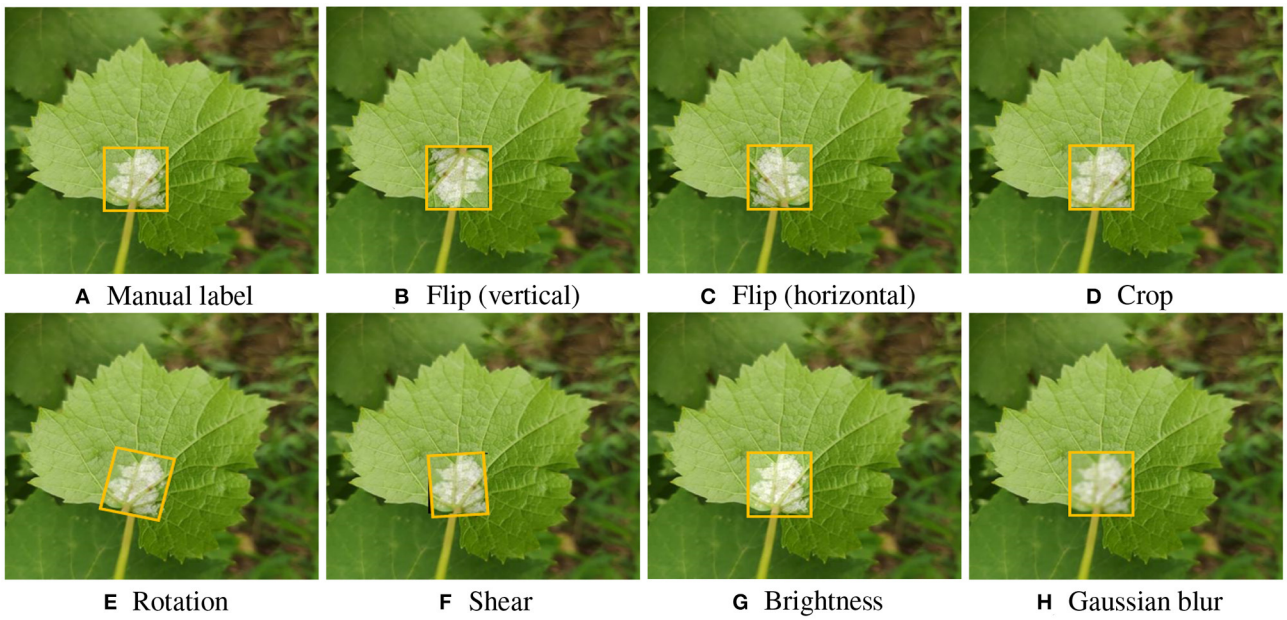


FIGURE 6 | Examples of Bounding box based data augmentation. **(A)** Manual label, **(B)** Flip (vertical), **(C)** Flip (horizontal), **(D)** Crop, **(E)** Rotation, **(F)** Shear, **(G)** Brightness, and **(H)** Gaussian blur.

TABLE 3 | Comparison of different GDM methods.

Method	Precision (%)	Recall (%)	F_1 (%)	mAP@0.5 (%)
YOLOv5-CA	85.59	83.70	84.63	89.55
YOLOv5-CA (with data augmentation)	88.82	83.63	86.15	90.02

were used to train the YOLOv5-CA network, which forces neural nets to optimize hyperparameters and generate a high-robust model. Some augmented bounding boxes on grape leaves can be seen in **Figure 6**.

As illustrated in **Table 3**, the data augmentation based YOLOv5-CA detection achieved a precision of 88.82%, a recall of 83.63%, and an F_1-score of 86.15%, which is slightly higher than those without data augmentation. The data augmentation positively influences the model's performance by increasing the size of the dataset and mitigating the over-fitting. The overall improvements demonstrated that the data augmentation module is helpful in the GDM detection, enlarging model learning ability and significantly improving detection performance.

4. DISCUSSIONS

This study presents a deep learning-based pipeline for automatic GDM detection in the vineyard. The grape leaf images acquired directly from the plants under field conditions were used to verify our proposed YOLOv5-CA approach. According to our experimental results presented in **Table 1**, a precision of 85.59%, a recall of 83.70%, an F_1-score of 84.63%, and a mAP@0.5 of 89.55% with the inference speed of 58.82 frames per second (FPS) was obtained for GDM detection. The detection accuracy of the proposed YOLOv5-CA is superior to that of state-of-the-art methods such as Faster R-CNN, YOLOv4, and YOLOv5. This high accuracy demonstrates the effectiveness of YOLOv5-CA for GDM detection of grapevine leaf images taken under field conditions. There yield results show that it is feasible to model visual symptoms for automatic GDM detection using a combination of the YOLOv5 and the CA mechanism. The proposed YOLOv5-CA automatically finds complex features capable of differentiating leaves with downy mildew symptoms and without any, which provides a precise and effective method for automatic disease detection.

On the other hand, the results presented in **Table 2** reveal the appropriate network input size in our experiments is 416×416. Additionally, **Table 3** compared the GDM detection performance with and without data augmentation, it shows that data augmentation enhances the GDM detection performance. The possible reason is that data augmentation increases the size of the dataset and brings more diversity to leverage the model training.

Although this study mainly focuses on GDM detection, it is suitable for multi-diseases detection (e.g., black spot, powdery mildew) after the model was re-trained with the dataset containing these diseases. As our approach uses RGB images, it would be a restriction for detecting GDM in the very earlier stage (i.e., non-visible symptoms) detection. However, if the multi-spectral images were acquired and used, our proposed YOLOv5-CA could be a potential tool to distinguish downy mildew from other leaf diseases/damage.

5. CONCLUSIONS AND FUTURE STUDY

To achieve an accurate and real-time intelligent detection of GDM under natural environments, an automatic YOLOv5-CA based detection method was proposed in this study. By combing YOLOv5 and coordinate attention, the GDM related visual features are well focused on and extracted, which boosts the GDM detection performance. Our proposed YOLOv5-CA achieved 85.59% detection precision, 89.55% mAP@0.5 with 58.82 FPS, which outperformed Faster R-CNN, YOLOv4, and YOLOv5. Moreover, the test results showed that the different disease levels of GDM and the illumination influence would not have a great impact on the GDM detection results, indicating the proposed method is feasible for the rapid and accurate detection of GDM. Ablation studies show that a network input size of 416×416 is favorable for fast GDM detection, and bounding box-based data augmentation boosts the GDM detection precision by 3.23%. The results exposed in this work indicate that downy mildew in grapevine can be automatically evaluated using artificial intelligence technology.

Overall, our approach achieved a good trade-off between speed and accuracy for GDM, and can be adapted to applications with autonomous-based smart farming. For future study, the multi-spectral information and edge computing will be exploited to further improve detection performance and computational efficiency.

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

ZZ: investigation, methodology, writing-review, and editing. YQ: data curation, methodology, formal analysis, and writing-original draft. YG: writing-review and editing. DH: resources and article revising. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Key Research and Development Program of China (2019YFD1002500), Ningxia Hui Autonomous Region Key Research and Development Program (2021BEF02015), and Ningxia Hui Autonomous Region Flexible Introduction Team Project (2020RXTDLX08).

REFERENCES

Abbas, A., Jain, S., Gour, M., and Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with c-gan synthetic images. *Comput. Electron. Agric.* 187, 1106279. doi: 10.1016/j.compag.2021.106279

Abdelghafour, F., Keresztes, B., Germain, C., and Da Costa, J.-P. (2020). In field detection of downy mildew symptoms with proximal colour imaging. *Sensors* 20, 4380. doi: 10.3390/s20164380

Adeel, A., Khan, M. A., Sharif, M., Azam, F., Shah, J. H., Umer, T., et al. (2019). Diagnosis and recognition of grape leaf diseases: An automated system based on a novel saliency approach and canonical correlation analysis based multiple features fusion. *Sustainable Comput.* 24, 1100349. doi: 10.1016/j.suscom.2019.08.002

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., and Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry* 11, 939. doi: 10.3390/sym11070939

Atanassova, S., Nikolov, P., Valchev, N., Masheva, S., and Yorgov, D. (2019). Early detection of powdery mildew (podosphaera xanthii) on cucumber leaves based on visible and near-infrared spectroscopy. *AIP Conf. Proc.* 2075, 160014. doi: 10.1063/1.5091341

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

Chen, M., Brun, F., Raynal, M., and Makowski, D. (2020). Forecasting severe grape downy mildew attacks using machine learning. *PLoS ONE* 15, e0230254. doi: 10.1371/journal.pone.0230254

Choi, H. C., and Hsiao, T.-C. (2021). "Image classification of cassava leaf disease based on residual network," in *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)* (Tainan: IEEE), 185–186.

Chorowski, J., Bahdanau, D., Serdyuk, D., Cho, K., and Bengio, Y. (2015). Attention-based models for speech recognition. *arXiv preprint* arXiv:1506.07503. doi: 10.48550/arXiv.1506.07503

Cristin, R., Kumar, B. S., Priya, C., and Karthick, K. (2020). Deep neural network based rider-cuckoo search algorithm for plant disease detection. *Artif. Intell. Rev.* 53, 4993–5018. doi: 10.1007/s10462-020-09813-w

Dinata, M. I., Nugroho, S. M. S., and Rachmadi, R. F. (2021). "Classification of strawberry plant diseases with leaf image using CNN," in *2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST)* (Yogyakarta: IEEE), 68–72.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 3111–3318. doi: 10.1016/j.compag.2018.01.009

Gutiérrez, S., Hernández, I., Ceballos, S., Barrio, I., Díez-Navajas, A. M., and Tardaguila, J. (2021). Deep learning for the differentiation of downy mildew and spider mite in grapevine under field conditions. *Comput. Electron. Agric.* 182, 1105991. doi: 10.1016/j.compag.2021.105991

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904–1916. doi: 10.1109/TPAMI.2015.2389824

Hernández, I., Gutiérrez, S., Ceballos, S., Iñíguez, R., Barrio, I., and Tardaguila, J. (2021). Artificial intelligence and novel sensing technologies for assessing downy mildew in grapevine. *Horticulturae* 7, 103. doi: 10.3390/horticulturae7050103

Hernández, S., and Lopez, J. L. (2020). Uncertainty quantification for plant disease detection using bayesian deep learning. *Appl. Soft. Comput.* 96, 1106597. doi: 10.1016/j.asoc.2020.106597

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* (Nashville, TN: IEEE), 13713–13722.

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: IEEE), 7132–7141.

Ji, M., Zhang, L., and Wu, Q. (2020). Automatic grape leaf diseases identification via unitedmodel based on multiple convolutional neural networks. *Inf. Process. Agric.* 7, 418–426. doi: 10.1016/j.inpa.2019.10.003

Kerkech, M., Hafiane, A., and Canals, R. (2020). Vine disease detection in uav multispectral images using optimized image registration and deep learning segmentation approach. *Comput. Electron. Agric.* 174, 1105446. doi: 10.1016/j.compag.2020.105446

Li, W., Liu, K., Zhang, L., and Cheng, F. (2020). Object detection based on an adaptive attention mechanism. *Sci Rep.* 10, 1–13. doi: 10.1038/s41598-020-67529-x

Liu, B., Ding, Z., Tian, L., He, D., Li, S., and Wang, H. (2020). Grape leaf disease identification using improved deep convolutional

neural networks. *Front. Plant. Sci.* 11, 11082. doi: 10.3389/fpls.2020.01082

Liu, E., Gold, K. M., Combs, D., Cadle-Davidson, L., and Jiang, Y. (2021). "Deep learning-based autonomous downy mildew detection and severity estimation in vineyards," in *2021 ASABE Annual International Virtual Meeting* (American Society of Agricultural and Biological Engineers).

Liu, J., and Wang, X. (2020). Tomato diseases and pests detection based on improved yolo v3 convolutional neural network. *Front. Plant Sci.* 11, 8198. doi: 10.3389/fpls.2020.00898

Liu, J., and Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. *Plant Methods* 17, 1–18. doi: 10.1186/s13007-021-00722-9

Liu, R., Cheng, Z., Zhang, L., and Li, J. (2019). Remote sensing image change detection based on information transmission and attention mechanism. *IEEE Access* 7, 1156349–1156359. doi: 10.1109/ACCESS.2019.2947286

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 8759–8768.

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

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, 558126. doi: 10.3389/fpls.2020.558126

Mohammadpoor, M., Nooghabi, M. G., and Ahmedi, Z. (2020). An intelligent technique for grape fanleaf virus detection. *Int. J. Interact. Multim. Artif. Intell.* 6, 62–67. doi: 10.9781/ijimai.2020.02.001

Mutka, A. M., and Bart, R. S. (2015). Image-based phenotyping of plant disease symptoms. *Front. Plant Sci.* 5, 7134. doi: 10.3389/fpls.2014.00734

Nagaraju, M., and Chawla, P. (2020). Systematic review of deep learning techniques in plant disease detection. *Int. J. Syst Assurance Eng. Manag.* 11, 547–560. doi: 10.1007/s13198-020-00972-1

Niu, Z., Zhong, G., and Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing* 452, 418–462. doi: 10.1016/j.neucom.2021.03.091

Qiao, Y., Kong, H., Clark, C., Lomax, S., Su, D., Eiffert, S., et al. (2021). Intelligent perception-based cattle lameness detection and behaviour recognition: a review. *Animals* 11, 3033. doi: 10.3390/ani11113033

Qiao, Y., Truman, M., and Sukkarieh, S. (2019). Cattle segmentation and contour extraction based on mask r-cnn for precision livestock farming. *Comput. Electron. Agric.* 165, 1104958. doi: 10.1016/j.compag.2019.104958

Ravi, V., Acharya, V., and Pham, T. D. (2021). Attention deep learning-based large-scale learning classifier for cassava leaf disease classification. *Expert. Syst.* 39, e12862. doi: 10.1111/exsy.12862

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* (Las Vegas, NV: IEEE), 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28, 911–999. doi: 10.48550/arXiv.1506.01497

Roy, A. M., and Bhaduri, J. (2021). A deep learning enabled multi-class plant disease detection model based on computer vision. *AI* 2, 413–428. doi: 10.3390/ai2030026

Singh, V., Sharma, N., and Singh, S. (2020). A review of imaging techniques for plant disease detection. *Artif Intell Agric.* 4, 229–242. doi: 10.1016/j.aiia.2020.10.002

Su, D., Kong, H., Qiao, Y., and Sukkarieh, S. (2021). Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Comput. Electron. Agric.* 190, 1106418. doi: 10.1016/j.compag.2021.106418

Tang, Z., Yang, J., Li, Z., and Qi, F. (2020). Grape disease image classification based on lightweight convolution neural networks and channelwise attention. *Comput. Electron. Agric.* 178, 1105735. doi: 10.1016/j.compag.2020.105735

Thet, K. Z., Htwe, K. K., and Thein, M. M. (2020). "Grape leaf diseases classification using convolutional neural network," in *2020 International Conference on Advanced Information Technologies (ICAIT)* (Yangon: IEEE), 147–152.

Tzutalin (2015). *Labelimg. git code (2015)*. Available online at: https://github.com/tzutalin/labelImg

Vishnoi, V. K., Kumar, K., and Kumar, B. (2021). Plant disease detection using computational intelligence and image processing. *J. Plant Dis. Protect.* 128, 19–53. doi: 10.1007/s41348-020-00368-0

Waghmare, H., Kokare, R., and Dandawate, Y. (2016). "Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system," in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)* (Noida: IEEE), 513–518.

Wang, J., Yang, J., Yu, L., Dong, H., Yun, K., and Wang, Y. (2021a). Dba_ssd: a novel end-to-end object detection using deep attention module for helping smart device with vegetable and fruit leaf plant disease detection. *Information* 12, 474. doi: 10.21203/rs.3.rs-166579/v1

Wang, P., Niu, T., Mao, Y., Liu, B., Yang, S., He, D., et al. (2021b). Fine-grained grape leaf diseases recognition method based on improved lightweight attention network. *Front. Plant Sci.* 12, 738042. doi: 10.3389/fpls.2021.738042

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), 3–19.

Yu, H.-J., and Son, C.-H. (2020). "Leaf spot attention network for apple leaf disease identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (Seattle, WA: IEEE), 52–53.

Zhang, K., Wu, Q., and Chen, Y. (2021). Detecting soybean leaf disease from synthetic image using multi-feature fusion faster r-cnn. *Comput. Electron. Agric.* 183, 1106064. doi: 10.1016/j.compag.2021.106064

Zhang, N., Yang, G., Pan, Y., Yang, X., Chen, L., and Zhao, C. (2020). A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sens.* 12, 3188. doi: 10.3390/rs12193188

Zhang, S., Zhang, S., Zhang, C., Wang, X., and Shi, Y. (2019a). Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Comput. Electron. Agric.* 162, 4122–4430. doi: 10.1016/j.compag.2019.03.012

Zhang, Y., Chen, Y., Huang, C., and Gao, M. (2019b). Object detection network based on feature fusion and attention mechanism. *Future Internet* 11, 9. doi: 10.3390/fi11010009

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Munich), 286–301.

Zhao, S., Peng, Y., Liu, J., and Wu, S. (2021). Tomato leaf disease diagnosis based on improved convolution neural network by attention module. *Agriculture* 11, 651. doi: 10.3390/agriculture11070651

Zhou, C., Zhang, Z., Zhou, S., Xing, J., Wu, Q., and Song, J. (2021). Grape leaf spot identification under limited samples by fine grained-gan. *IEEE Access* 9, 1100480–1100489. doi: 10.1109/ACCESS.2021.3097050

Zhu, J., Wu, A., Wang, X., and Zhang, H. (2020). Identification of grape diseases using image analysis and bp neural networks. *Multimed Tools Appl.* 79, 14539–14551. doi: 10.1007/s11042-018-7092-0

Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., and Le, Q. V. (2020). "Learning data augmentation strategies for object detection," in *European Conference on Computer Vision* (Springer), 566–583.

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.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 August 2022
DOI 10.3389/fpls.2022.892388

OPEN ACCESS

EDITED BY
Yongliang Qiao,
The University of Sydney, Australia

REVIEWED BY
Fuyang Tian,
Shandong Agricultural University, China
Xuewei Zhang,
Xi'an Technological University, China
Rashad Hegazy,
Kafrelsheikh University, Egypt

*CORRESPONDENCE
Jizhan Liu
1000002048@ujs.edu.cn

SPECIALTY SECTION
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

RECEIVED 09 March 2022
ACCEPTED 01 July 2022
PUBLISHED 04 August 2022

CITATION
Xie B, Liu J, Jiang H, Cai L, Liu L and Li Y (2022) Development and experimental analysis of a small integrated edge navigation sensor based on principle of circular arc array. *Front. Plant Sci.* 13:892388. doi: 10.3389/fpls.2022.892388

Development and experimental analysis of a small integrated edge navigation sensor based on principle of circular arc array

Binbin Xie[1], Jizhan Liu[1]*, Houkang Jiang[1], Lianjiang Cai[1], Lu Liu[2] and Yuanxiang Li[3]

[1]Key Laboratory of Modern Agricultural Equipment and Technology, Jiangsu University, Zhenjiang, China, [2]Institute of Technology, Anhui Agricultural University, Hefei, China, [3]Intelligent Equipment Company of Hunan Xiangyuan Golden Spike, Loudi, China

Signal, accuracy, and real-time performance of satellite, radar, and machine vision is a subject of concern in various complex agricultural environments. Therefore, the demand for a robust navigation sensor for indoor and vertical agricultural environment remains crucial, and it is a significant subject. In view of this, the relative edge pose detection method based on the ideal target band principle of the lateral center arc array, in this research, a small integrated arc array navigation sensor module based on adaptive detection arc technology, is developed, which costs about $100, autonomous edge navigation position, and attitude detection is realized in facility agriculture environment with continuous structured corridor or roadside features. In this research, a coupling method of reducing the radius of distance sensor arrangement, adjusting the unequal center angle, and increasing the detection distance is used to realize the miniaturization of the arc array arrangement. A semicircular modular rocket was designed to slide and adjust the center angle of the distance sensor, and the longitudinal installation position of the modularized sensor was adjusted by translating the circular arc of the detection; the convenient moving arrangement under different vehicle width and detection arc characteristics is realized. An adaptive construction method of detecting a circular arc based on self-calibrating detection distance of a distance sensor is proposed, which effectively reduces the difficulty of arranging the lateral central circular arc array; the fast construction of lateral detection arc is realized. In addition, in order to improve the accuracy and stability of the pose detection, the Mahalanobis distance algorithm and the standard Kalman filter are used to optimize the estimation of the ranging information and the relative pose of the edge. The experimental results show that the small integrated arc array navigation sensor module can independently construct photoelectric detection arcs with different characteristics to detect the position and attitude of the relative edge. When the road surface is concave and convex, the small integrated arc array navigation sensor module can still maintain the stable position and attitude detection of the relative edge for more than 30 s. In addition, when the walking speed of the autonomous navigation

platform is 0.15 m/s to 0.35 m/s, the detection errors of lateral deviation and heading deviation relative to the road edge are less than 40 mm and 4.5°, respectively. The small integrated arc array navigation sensor module is less affected by the change of operating speed, and still has good accuracy and stability. The results show that the modularized edge navigation sensor has the advantages of fast and convenient use, high accuracy, and low cost; it can be applied to autonomous edge navigation control in greenhouse and plant and animal factories.

KEYWORDS

facility agriculture, corridor environment, edge walking, navigation sensors, miniaturization, integration, modularization

Introduction

Motivation and background

Facility agriculture planting makes crops unaffected by geographical environment, climate, and seasonal changes, guarantees crop yield, saves planting costs, and improves economic income benefits. Therefore, facility agriculture has received extensive attention and high-tech applications in countries around the world (Ferentinos et al., 2017; Soiket et al., 2019). Under the background that the world's facility agriculture continues to develop toward high and new technologies, such as informatization, intelligence, and automation, by 2021, the total area of global facility agriculture has exceeded 4.6 million hm^2, facility agriculture has developed rapidly, and the total planting area has continued to grow rapidly in China; the scale of facility agriculture has exceeded 4 million hm^2, accounting for 85% of the global total. The autonomous navigation operation platform in the agricultural facility environment is mainly responsible for a large number of mobile operations, such as frequent handling, spraying, transplanting, harvesting, etc. Due to the complex, changeable, and more interference agricultural environment of the facility, mainstream autonomous navigation methods, such as satellite, radar, and machine vision, have the problems of weak signals, poor accuracy, and lagged real-time, which are difficult to meet operational requirements. Therefore, combining the characteristics of agricultural scenes, researching intelligent navigation technology and special sensors suitable for the facility agricultural environment is the key to realizing efficient autonomous navigation operations, and it is also a necessary condition for the development of unmanned facility agriculture (Liang et al., 2018; Soiket et al., 2019).

Autonomous navigation is the basis for the implementation of unmanned precision agriculture, which can effectively reduce labor intensity and improve operation accuracy and efficiency. In recent years, the rapid development of satellite autonomous navigation technology has been widely used in the field of agricultural environment. For example, Han et al. (2020) developed an agricultural machinery automatic driving system with multi-sensor data fusion algorithm based on GNSS, RTK, and motion sensors, etc.; the root mean square error of the path following navigation test is less than 9 cm. Takai et al. (2014) designed an autonomous navigation system for agricultural machinery with adaptive control algorithm based on RTK-GPS and inertial sensors, and the lateral deviation control error was less than 5 cm. Li et al. (2019) used the Kalman filter algorithm to fuse information from sensors, such as GPS, gyroscope, and electronic compass, to obtain accurate navigation data, and carried out experimental tests on the Tieniu 654 tractor and Lovol TG1254 tractor produced in China. Yunpeng et al. (2019) developed an automatic navigation system for agricultural machinery based on real-time dynamic positioning technology, and differential GPS technology, which can achieve continuous high-precision positioning and navigation under the condition of GPS signal stability. With the rapid development of differential RTK technology, satellite positioning technology represented by GPS and BDS has been widely used in agriculture, road transportation, and ship transportation for precise navigation, but the indoor environment of facility agriculture will seriously block satellite signals. The signal is discontinuous, and the satellite navigation fails.

In order to realize autonomous navigation in unstructured agricultural environment, radar and machine vision autonomous navigation technology has become the main research direction of agricultural autonomous navigation technology. For example, Jia et al. (2015) realized the edge detection of a flat greenhouse road based on a two-dimensional lidar, which has a poor adaptability to the flat slope and uneven road on both sides of the road. Wang et al. (2012) used machine vision to extract the sensitive area of the heating tube in the greenhouse ridge-planted tomato environment, and realized the autonomous navigation through the center baseline fitting between the ridges. Gao and Ming (2014) used K-means algorithm and the morphological erosion method to cluster and segment the collected greenhouse environment images, eliminate redundant interference information, and reduce

the impact of illumination on machine vision recognition navigation paths. In the orchard tree row environment, corn and other high-stalk crops are planted in rows; lidar is often used to detect crop stalks to complete the line center baseline fitting and to achieve autonomous path perception and navigation in unstructured agricultural environments (Hiremath et al., 2014; Malavazi et al., 2018). In order to solve the problem of incorrectly fitting the navigation path and identifying obstacles in the unstructured environment where the distribution of path features is seriously uneven, the image features of the operating environment are recognized by machine vision to improve the accuracy of the navigation path fitting in complex environments, and can achieve rapid identification and classification of static and dynamic obstacles to improve autonomous navigation safety in unstructured environments (Yang et al., 2018; Inoue et al., 2019; Xu et al., 2021). Although radar and machine vision autonomous navigation technology has carried out many studies and achieved certain results in complex unstructured greenhouse environment and semi-natural interference conditions, such methods need to establish complex algorithms and are rich in information in facility agricultural environment. It is easy to be disturbed during the operation, and has not yet reached the point of application in actual production.

Recently, facility agriculture uses engineering technology and industrial production methods to provide a good growth environment for crops or animals. The environment of standardized greenhouse and plant-animal factory is usually equipped with furrows and cultivation troughs to form three-dimensional cultivation devices, presenting the characteristics of continuous structured corridors or roadsides, as shown in **Figure 1**. During the cultivation of fruits, vegetables, flowers and other plants, as well as the breeding of poultry and other animals, numbers of automatic operations, such as transporting, applying chemicals, transplanting and harvesting, should be accomplished by autonomous navigation platform in the corridor. So it is the current trend to use corridor or road-edge features to conduct curb line navigation technology research according to these features. The curbside characteristics of the environment of facility agriculture corridor, as shown in **Figure 2**. So, it is the current trend to use corridor or road edge features to conduct curb line navigation technology research according to these features. The curbside characteristics of the environment of facility agriculture corridor, as shown in **Figure 2**. The premise of autonomous edge navigation is to obtain the position information of the autonomous navigation platform along the relative road by using photoelectric switch. The photoelectric switch was installed on the side of the autonomous mobile platform walking along the edge, and judge the relative relationship between the autonomous mobile platform and the road edge based on the high-level and low-level changes of the photoelectric switch. In order to realize the detection of different road edge features, Ju (2017) adjusted the installation inclination of the photoelectric switch according to **Figure 3**, in combination with the geometric relationship between the convex edge and the sunken edge in **Figure 2**. During the operation of the autonomous mobile platform, when the horizontal distance to the curb is greater than or less than the safe distance, the high-level and low-level triggering states of the photoelectric switch will be changed to achieve real-time detection of the curb.

The realization of distance or collision perception based on mechanical contact, infrared photoelectric or ultrasonic sensors has been applied on many hair surfaces such as home and social service robots. However, they can only complete edge anti-collision or anti-drop control, and cannot accurately determine the relative position and posture of the autonomous navigation platform and edge, and cannot meet the requirements of high precision operation in facility agriculture (Zhao et al., 2012; Wang and Liang, 2015). In order to solve the above problems, many researchers have carried out the design of autonomous curb line navigation methods based on the road edge characteristics of the facility corridor environment. An ultrasonic or infrared photoelectric sensor is arranged laterally on the autonomous navigation platform to complete the distance measurement of the road edge to avoid collision, but the inclination state of the relative wall cannot be calculated and judged, causing the front or rear of the autonomous mobile platform to easily collide with the road edge, which is difficult to guarantee safe, efficient, and reliable autonomous driving along the road edge (Feng et al., 2012; Zhou, 2014).

For the problem that the single ultrasonic or infrared photoelectric sensor cannot obtain the heading deviation and transverse deviation accurately, multiple photoelectric switches are arranged in front of the autonomous mobile platform to measure the distance along the road, and multiple distance signals are used to construct fuzzy logic fusion to complete the curb line navigation. However, the calculation of this method is complicated, and the experimental results show that the walking posture is not stable enough (Xu et al., 2010; Yuan and Li, 2013). Du (2010) arranged several photoelectric switches in parallel lines on the side of the autonomous navigation platform to achieve autonomous walking along the road edge, but this method can only achieve a rough determination of the deviation state, and cannot complete the measurement of specific position and orientation, so it is difficult to ensure the smoothness of the curb line navigation and control accuracy.

In order to improve the smoothness and control accuracy of edge navigation, Ju (2017) analyzed the arrangement of a photoelectric switch array, as shown in **Figure 4**. As shown in **Figure 4A**, due to the fixed lateral position of the photoelectric switch, the local horizontal discrete distance of the mobile platform relative to the road edge can be obtained from the photoelectric switch state. But, in the same photoelectric switch state, the heading angle of the mobile platform can change greatly, so the effective heading angle (γ) cannot be

FIGURE 1
Modern standardized facility agricultural corridor environment.

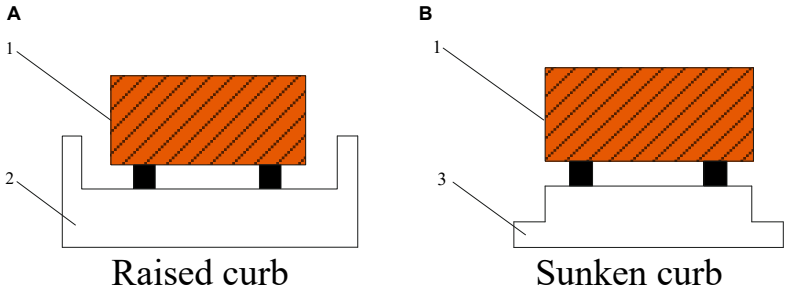


FIGURE 2
Characteristics of two types of curbs: (1) An autonomous moving platform, (2) A raised curb, and (3) A sinking curb.

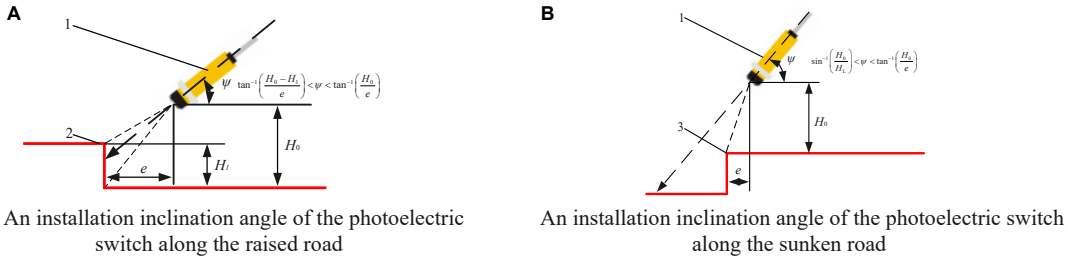


FIGURE 3
A layout principle of a photoelectric switch to identify the road edge: (1) A photoelectric switch, (2) A raised road edge, (3) A sunken road edge, *e*. A safety distance between the autonomous navigation platform and the curb line during driving (mm), H_0. Height of the photoelectric switch installed on the mobile platform (mm), H_1. Height of the raised curb (mm), H_L. A range set by the photoelectric switch (mm), ψ. An inclination angle between a photoelectric switch and a road edge (°).

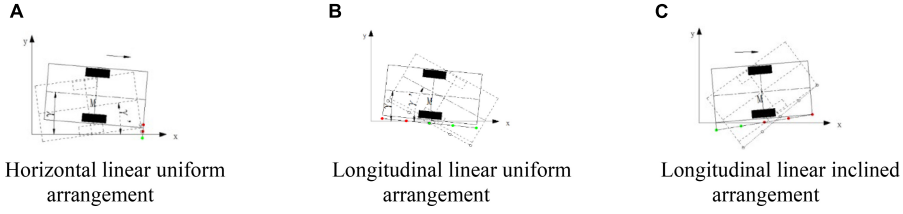


FIGURE 4
A linear arrangement scheme of a photoelectric switch array: γ. The heading angle of the autonomous mobile platform relative to the road along.

directly obtained from the current state of the photoelectric switch group. The information can only be obtained from the installation position of the photoelectric switch group, the time of state change before and after the photoelectric switch group, and the motion state of the mobile platform, and then derive the heading angle (γ) of the mobile platform through a certain algorithm, so it is difficult to provide more complete and effective information for the curb line navigation control of the autonomous mobile platform. As shown in **Figure 4B**, the horizontal discrete distance of the mobile platform relative

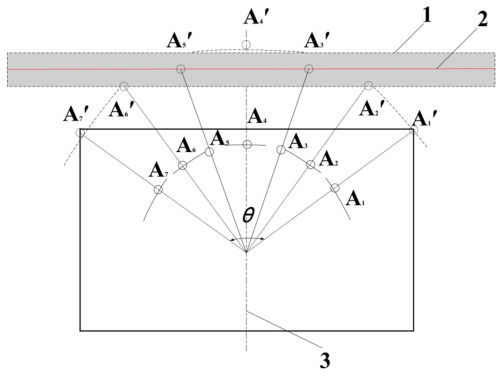


FIGURE 5
Seven principle of autonomous curb line navigation of a photoelectric switch arc array: (1) An ideal target belt, (2) A road edge, (3) An autonomous mobile platform, A_1∼A_7. A photoelectric switch, A_1'∼A_7'. A detection point of photoelectric switch, θ. The central angle of the photoelectric switch arc array.

to the road edge can be obtained, but it is difficult to obtain a more accurate heading angle, which cannot provide complete and effective information for the control of the autonomous mobile platform's curb line navigation, As shown in **Figure 4C**, the horizontal discrete distance of the mobile platform relative to the road edge can also be obtained, but it is also difficult to obtain a more accurate heading angle, which cannot provide complete and effective information for the control of curb line navigation. To sum up, it is difficult to obtain the precise lateral deviation and heading deviation of an autonomous mobile platform relative to the road edge at the same time with the linear arrangement of photoelectric switches.

In order to solve the problem that the precision lateral deviation and the course deviation from the autonomous mobile platform cannot be obtained with the linear arrangement of photoelectric switches, Ju et al. (2017) proposed an autonomous curb line navigation method based on seven photoelectric switches, took the trigger number of a photoelectric arc array signal and the number of the trigger center as dual indicators. The position and orientation detection model is established, and divides nine relative road edge position and orientation states according to the values of the dual indicators, and uses the fuzzy control method to realize real-time control of curb line navigation. As shown in **Figure 5**, under the detection arc of the photoelectric switch array set in the test, the control accuracy of the lateral deviation and the heading deviation between the autonomous mobile platform and the road edge were stable within the range of -35 mm∼15 mm and ±5°, respectively. The experimental results confirmed the feasibility of the method. However, this method needs to install and debug the photoelectric switch according to the actual vehicle body length and navigation accuracy requirements before use, which makes the use process cumbersome, and has not been deduced and verified for generalization and small integration, and has not yet entered the actual production operation. Therefore, building a small and integrated photoelectric switch arc array module can bring many conveniences into production practice, and can promote the development of low-cost and fast curb line navigation technology.

In response to this problem, this research developed a small integrated photoelectric arc array switch navigation sensor for facility corridor environment. The central photoelectric arc array based on the principle of the ideal target band, the system establishes a real-time position and an orientation detection method with unknown number of sensors. By adjusting the vertical installation position of the modular sensor through the shift detection arc method, the convenient mobile layout under different vehicle widths and detection arc characteristics can be achieved, which effectively reduces the difficulty of the lateral central arc array placement. Therefore, the small integrated circular arc array navigation sensor module can be widely used in many autonomous navigation platforms in the market, and can provide high-precision position detection. Specifically, the main features of the sensor are that it can realize convenient and fast curb line navigation with low cost and high precision. Specifically, the main features of the sensor are that it can realize convenient and fast curb line navigation with low cost and high precision.

Scope and contribution

The overall goal of this research is to develop an intelligent navigation sensor suitable for the facility agricultural environment. Based on the ideal target principle of the lateral center arc array, we propose a relative edge pose detection method. We have developed a small integrated edge navigation module that adaptively constructs and detects arcs, and can realize low cost and fast edge navigation in the a facility agricultural corridor environment. The main contributions of this paper are as follows:

(1) In the early stage, it was verified the feasibility of a circular array edge-navigation method. In this paper, the edge pose state judgment rules and the pose deviation calculation method of an unknown sensor number were established, which can realize the autonomous edge pose detection under the configuration of different sensor numbers, arc center angles, and arc radiuses, and improve the universality of the arc array edge navigation method.
(2) In this article, a miniaturization scheme is proposed to reduce the sensor layout radius, adjust the non-equidistant center angle, and increase the detection distance. Under the condition of accurately reflecting the relative pose relationship between the head and the tail of the autonomous navigation platform and the road

edge, and ensuring the resolution of edge pose detection and recognition, we realized the miniaturization of the arc array sensors.

(3) In order to realize the rapid adjustment and convenient arrangement of the navigation sensors along the edge, we propose an adaptive construction method of the detection arc based on the detection distance of the self-calibration sensor. According to geometric composition of the detection arc, a self-calibration equation of detection distance of sensors for each position and a calculation program of rapid arrangement of the arc center angle are established. The longitudinal installation position of the modularized sensor is adjusted by the method of the translation detection arc to match the installation position under different vehicle widths and detection arc characteristics.

(4) In order to improve the accuracy and stability of a relative edge position pose and enhance the performance of restraining the interference of environmental factors, we used the improved Mahalanobis distance algorithm to eliminate the abnormal ranging noise and environmental sudden change interference, and use the Kalman filter algorithm to optimally estimate the relative edge position and attitude state so as to achieve the purpose of weakening the relative edge position and attitude state detection error.

(5) Finally, we developed a low-cost small integrated arc array edge navigation sensor module by using embedded microcontroller and carried out experimental verification.

Article structure

The rest of the research is organized as follows: the second part introduces the design scheme of the miniaturized optimized integrated arc array edge navigation sensor proposed in this study. The third part introduces the system design of a small optimized integrated arc array edge navigation sensor. The test design and results are presented in section "Experiment." Finally, a summary of the study is provided in section "Results."

Materials and methods

In this section, we successively introduce the design architecture of the arc array edge navigation sensor module, the arc array edge navigation principle and universal modeling, the photoelectric arc array miniaturization integration scheme, the arc array adaptive self-calibration construction method, and the relative edge position and orientation detection error reduction method.

Design architecture of a small integrated arc array edge navigation sensor module

The essence of edge navigation is that the controller detects the relative pose deviation between the autonomous mobile platform and the continuous structured corridor or curb features in real time to correct the path. In order to solve the problems of a weak signal, poor accuracy and insufficient real-time performance of the current autonomous navigation technologies, such as satellite, radar, and machine vision in the facility agricultural environment, small integrated modules are used to achieve the purpose of accurate acquisition of the relative position and attitude along the edge of the autonomous mobile platform. In this paper, a general pose state detection model is built based on the relative edge pose detection method based on the ideal target band principle of the lateral central arc array in the author's (Ju et al., 2017) previous research. A small integrated edge navigation sensor model is built by using the coupling method of reducing the sensor layout radius, adjusting the center angle of the non-bisection circle, increasing the detection distance. The active self-calibration sensor detection distance method is used to adaptively construct the edge pose detection arc, and the relative edge pose detection accuracy is improved by improving the Mahalanobis distance algorithm and Kalman filter algorithm. As shown in **Figure 6**, the development of a small integrated edge navigation sensor based on the principle of arc array is realized.

The principle and universal modeling of arc array edge navigation

The principle of arc array edge navigation

The edge position and attitude detection of the arc array is to install an odd number of distance measuring sensors ($2N + 1$) on one side of the autonomous navigation platform in a centrosymmetric manner, and construct the edge position and attitude detection arc, as shown in **Figure 7**. The center position distance measuring sensor ($N + 1$, sensor detection point $A_{N+1}{}'$) is aligned with the horizontal centerline of the autonomous navigation platform. In addition, the arc array edge navigation takes the width formed between the connecting line between the sensor detection point $A_2{}'A_{2N}{}'$ and the detection point $A_{2N-1}{}'A_{2N+1}{}'$ as the ideal target band for position and attitude adjustment, and controls the autonomous navigation platform to always move smoothly within the attitude range relatively stable with the edge of the road. At the same time, we can see that the position distribution of the detection points of each sensor determines the characteristics of different detection arcs.

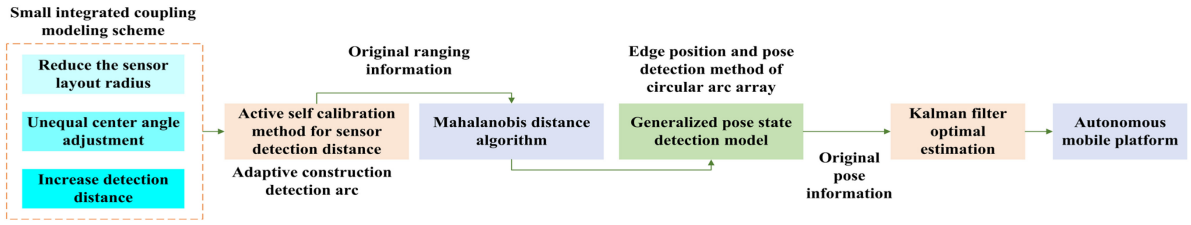


FIGURE 6
An overall architecture of a small integrated arc array edge navigation sensor module.

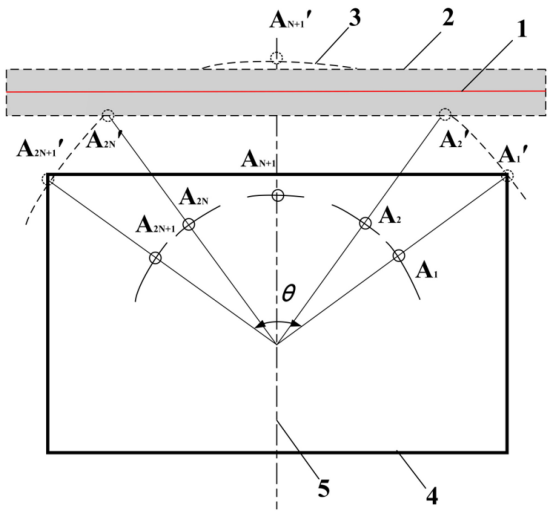


FIGURE 7
An autonomous edgewise navigation model of an N (odd) photoelectric switch lateral arc array: (1) A road edge, (2) An ideal target band, (3) A detection arc, (4) An autonomous mobile platform, (5) A transverse centerline.

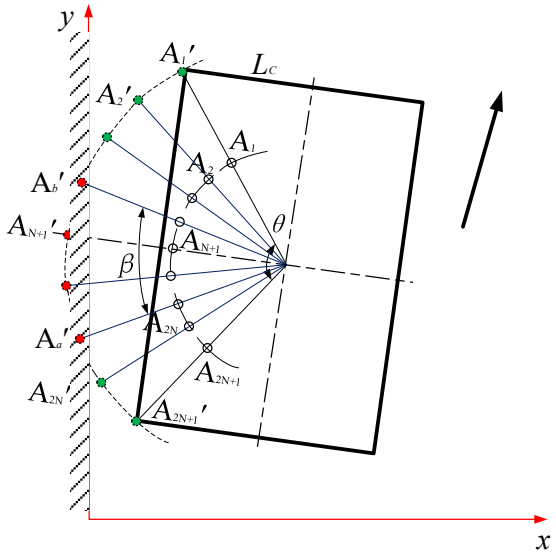


FIGURE 8
Calculation and analysis of pose deviation based on a photoelectric switch state.

TABLE 1 Generalized position and orientation state classification rules for unknown number of photoelectric switches.

Orientation	Position		
	$1 \leq N_d<3$	$N_d = 3$	$3<N_d \leq 2N-1$
$N<N_f<N+1$	A: Far away roadside to the outside yaw	B: Outside yaw	C: Near roadside to the outside
$N_f = N+1$	D: Far away roadside	E: Normal	F: Near roadside
$N+1<N_f<(4N+3)/2$	G: Far away roadside to the inside yaw	H: Inside yaw	I: Near roadside to the inside yaw

Position and pose detection along the edge of the generalized arc array

In order to realize the universal edge pose detection of the arc array, establish the geometric relationship when the autonomous navigation platform has position and attitude deviation; as shown in Figure 8, the green dot represents the non-triggered ranging sensor, and the red dot represents the triggered ranging sensor. The edge pose state detection and judgment rules of unknown sensor number are established according to the previous research of the author (Ju et al., 2017), as shown in formula (1). The tolerance range of the calculation error of the heading deviation and transverse deviation of the autonomous navigation platform relative to the road edge can be further simplified, as shown in Formula (2). As shown in Table 1, the two parameters, signal trigger number N_d and trigger center serial number N_f, are used to judge the relative pose state type of the autonomous navigation platform along the road. The larger the N_d is, the closer the autonomous navigation platform is to the road edge; on the contrary, the smaller the N_d is, the farther the autonomous navigation platform is from the road edge. N_f is the average value of the serial numbers

of the photoelectric switches that are triggered when the road edge is detected outside the range of the photoelectric switch. The larger the N_f is, the more the head of the autonomous navigation platform deviates from the road edge; on the contrary, the smaller the N_f, it means that the head of the autonomous navigation platform moves more toward the road edge.

$$\begin{cases} \tau = \frac{\theta}{2N} \\ N_d = b - a + 1 \\ N_f = \frac{b+a}{2} \\ \frac{(3+2N-b-a)\theta}{4N} < \gamma < \frac{(1+2N-b-a)\theta}{4N} \\ D = R\cos\frac{\beta}{2} - \left(R\cos\frac{\theta}{2} - L_C\right)\cos\gamma \\ \beta = \frac{(b-a+1)\theta}{2N} \end{cases} \quad (1)$$

In the formula: τ—The angle between adjacent sensor detection points; θ—The central angle of the circular arc; N—The total number of unilateral ranging sensors; a—Minimum serial number of the triggered ranging sensor; b—Maximum serial number of the range sensor; N_d—Number of triggered ranging sensors; N_f—Average value of serial number of triggered ranging sensor; γ—The heading deviation of the autonomous navigation platform relative to the road, °; D—Lateral deviation of the center point of the autonomous navigation platform relative to the road edge, mm; R—The detection radius of the photoelectric array switch, mm; L_C—An autonomous navigation platform width half, mm.

$$\begin{cases} \frac{(3+2N-2N_f)\theta}{4N} < \gamma < \frac{(1+2N-2N_f)\theta}{4N} \\ R\cos\frac{N_d\theta}{4N} - \left(R\cos\frac{\theta}{2} - L_C\right)\cos\frac{(2+2N-N_d)}{4N} < D < R\cos\frac{N_d\theta}{4N} \\ - \left(R\cos\frac{\theta}{2} - L_C\right)\cos\frac{(2N-N_d)}{4N} \end{cases} \quad (2)$$

A miniaturization scheme of the arc array edge navigation sensor

The relative edge position and orientation detection method based on the ideal target band principle of the lateral central arc array has been proved to be feasible by the author's previous research. However, the arc array occupies a large proportion of space and the installation and layout of sensors are cumbersome. At present, the small-scale integrated system has not been developed, and the user-friendliness and practicality are insufficient. In this section, we carry out small-scale integrated scheme design for the purpose of convenient use.

Analysis and determination of a miniaturization scheme

In essence, the miniaturization of the arc array is to reduce the plane layout size of the distance measuring sensor and realize the accurate acquisition of the position and attitude along the edge of the autonomous navigation platform. Through mathematical and geometric analysis, scaling the detection arc with an equal ratio column and scaling the detection arc with a variable center angle can reduce the layout size of the ranging sensor, as shown in **Figure 9**. The proportional column scaling arc in **Figure 9A** is to shrink A_{Li}' and A_{Ri}' inward to B_{Li}' and B_{Ri}' along the original radius direction without changing the center angle so as to reduce the arc radius and chord length of the proportional column so as to reduce the detection arc and achieve the purpose of miniaturization. When the longitudinal length of the autonomous mobile platform is large, the detection arc cannot reflect the relative position and orientation relationship between the head and the tail of the autonomous navigation platform and the road edge, resulting in a blind area along the curb line navigation, which cannot realize omni-directional detection and is not feasible. **Figure 9B** – the variable central angle scaling of the arc is to arrange the photoelectric switches A_{Li}' and A_{Ri}' on the original detection arc to the central position of the photoelectric switch under the condition that the radius of the detection arc remains unchanged. By changing the center-angle evenly distributed sensors, the layout chord length and the center angle of the photoelectric array are reduced so as to reduce the detection arc and achieve the purpose of miniaturization. The principle of this method is to move the detection point closer to the center of the detection arc, without changing the relative position and orientation relationship between the detection arc and the head and the tail of the autonomous navigation platform and the road edge, and there will be no blind area of curb line navigation.

According to **Figure 8**, it is found that the detection arc is formed by a series of curves connected by fixed points. In order to solve the problems in the sensor layout scheme of scaling, the detection arc with an equal ratio column and scaling the detection arc with a variable center angle, the two schemes are fused. The coupling adjustment layout method of reducing the ranging sensor layout radius - adjusting the unequal center angle - increasing the detection distance is adopted to realize the miniaturized layout design of the arc array, as shown in **Figure 10**. The semicircular shell with an arc chute (the green part in the figure) is designed. By adjusting the center angle and detection distance of different sensors, the original detection arc (the purple arc in the figure) is constructed by matching the detection points on the original detection arc to realize the miniaturization of the integrated arc array configuration. The pink part in **Figure 10** is the embedded system integrated shell, and the green line is the semicircular shell radius of the arc chute, the blue line is the distance from the detection point of each position sensor, the horizontal red line is to detect the arc chord length, and the vertical red line is to detect the distance

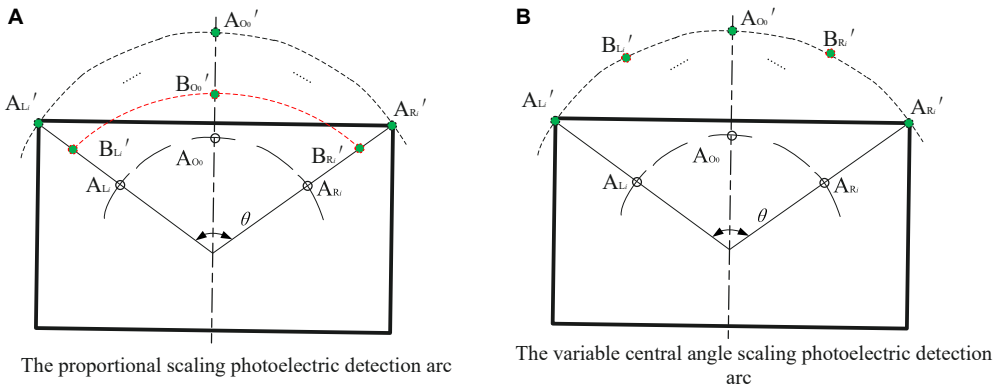


FIGURE 9
A schematic diagram of reducing the detection arc.

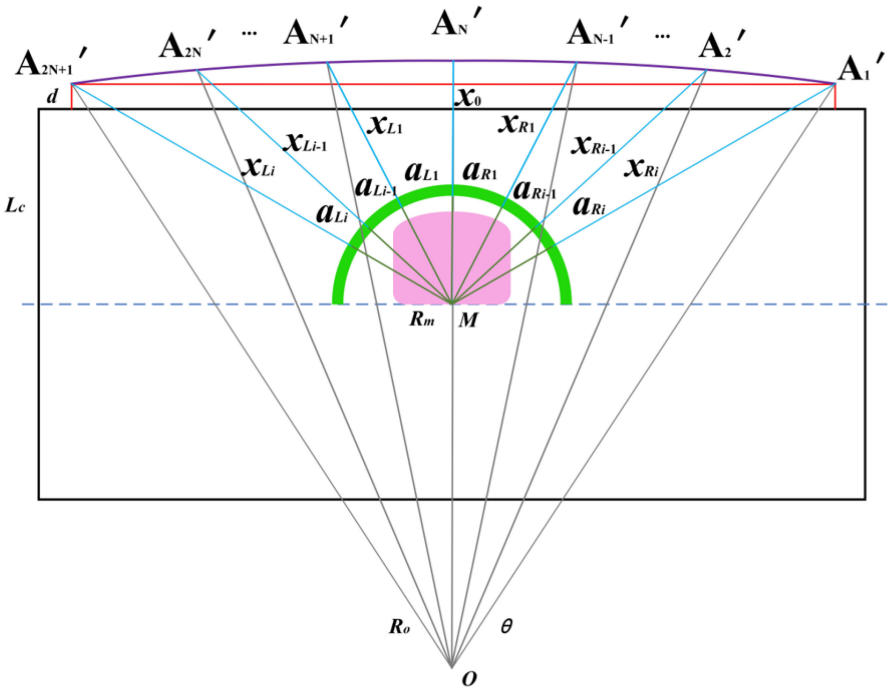


FIGURE 10
A coupling adjustment method of reducing the layout radius of ranging sensors, adjusting the center angle of non-equidistant circles and increasing the detection distance: x_0. Distance detected by sensors at center positions, x_{Li}. Distance detected by sensors at different positions on the left, x_{Ri}. Distance detected by sensors at different positions on the right, a_{Li}. The center angle of the sensor at different positions on the left, a_{Ri}. The center angle of the sensor at different positions on the right, R_m. A miniaturized arc array module original radius, R_o. An original detection arc radius, $A_1' \sim A_{2N+1}'$. An ultrasonic sensor detection point.

from the arc chord length to the side of the autonomous navigation platform.

Modular design of a circular arc array edge navigation sensor

According to the arc array miniaturization integration scheme of reducing the distance measuring sensor layout radius – non-equidistant center angle adjustment – increasing the detection distance in Figure 10, the arc array edge navigation sensor module is designed. As shown in Figure 11, it consists of an arc chute semicircular shell, an embedded system integration shell, a center-angle adjustment calibration scale plate, an angle-sliding adjustment plate, an upper cover plate, a power input port, a navigation signal output port, and an installation positioning hole. In order to realize the matching of sensor detection points, fix the ranging sensor on the angle sliding

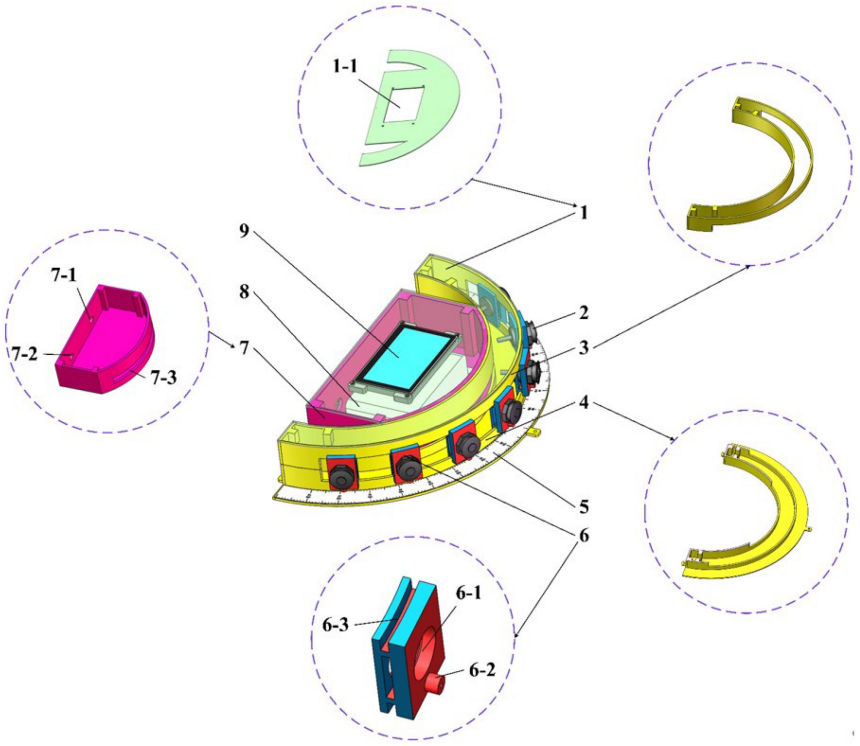


FIGURE 11
A structure of a circular arc array edge navigation sensor module: (1) An upper cover, (1-1) A man-machine interaction display mounting port, (2) A ranging sensor, (3) An arc chute half-round shell, upper part, (4) A lower part of an arc chute semicircular shell, (5) A calibration plate for central angle adjustment, (6) An angle-sliding adjustment plate, (6-1) A range sensor, (6-2) A bottom positioning bolt hole, (6-3) A chute, (7) Embedded System-integrated housing, (7-1) A power input port, (7-2) A navigation signal output port, (7-3) A range finder through a slot, (8) Embedded-controller housing, (9) A man-machine interaction display screen.

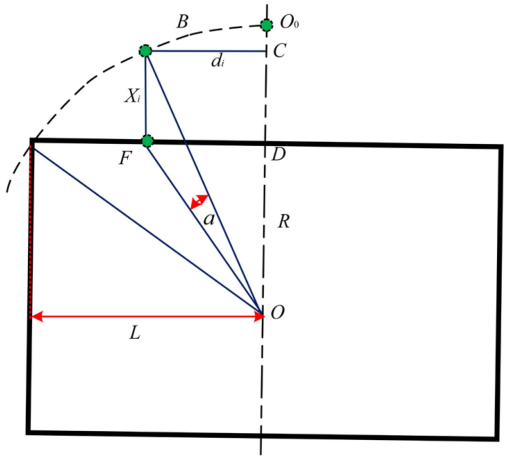


FIGURE 12
An analysis diagram of distance adjustment of a photoelectric switch detection point at each position.

adjusting plate to cooperate with the center angle adjusting calibration plate to slide freely to any angle alignment on the arc chute semicircular shell, lock the position of the sensor on the arc chute semicircular shell through the bottom positioning bolt, and adjust the sensor detection distance to complete the detection point setting. At the same time, the controller is installed in the embedded system integrated shell, and the arc surface of the shell is connected with the concave surface inside the semicircular shell of the arc chute to form a whole, which is conducive to the integrated sensor module design. In order to realize the human-computer interaction friendliness of the arc array edge navigation sensor module and facilitate the detection of arc parameters self-regulation, the upper cover plate with a serial port touch screen is designed, which can improve the power safety and water resistance. In addition, in order to reduce the interference of the power signal to the navigation signal transmission, the power line and the signal line are branched independently. The power input port and the navigation signal output port are placed at the tail of the embedded system-integrated shell and led out by the waterproof connector.

Modeling the detection distance and the central angle of sensors in different positions

It can be seen from Figure 11 that the arc array miniaturization integration scheme based on reducing the

layout radius of the ranging sensors - adjusting the unequal center angle – increasing the detection distance is to match the detection points of the sensors by adjusting the center angle and detection distance of the sensors at different positions to complete the detection arc structure. Therefore, the center angle layout and detection distance adjustment modeling are carried out for the ranging sensors at different positions. Since there is an odd number of ranging sensors on the edge navigation sensor module of the circular arc array, and the $N + 1$ sensors about the central position are symmetrically distributed and equal in number. Therefore, only one side needs to be considered when modeling the center angle arrangement and detection distance adjustment of distance sensors at different positions. As shown in **Figure 12**, the center angle arrangement and the detection distance adjustment model are established with different position sensors on the left. Point B is the original position of the detection point of the i-th ranging sensor, point F is the position of the i-th ranging sensor on the side of the autonomous navigation platform, and O_o is the central position ranging sensor. The cosine value of $\angle BOF$ of ΔBOF is solved by using the cosine formula of two-angle difference, and then the BF distance is solved by using the cosine theorem, as shown in Equation (3). Then, according to **Figure 10**, the detection distance of distance sensors at different positions is solved by using the Pythagorean theorem of the right triangle, and then the center angle of the layout of distance sensors at different positions is solved by using the cosine theorem, as shown in Equation (4).

$$\begin{cases} \theta_i = i \bullet \frac{\theta}{2N} \\ d_i = R_o\left(1 - \cos\frac{\theta}{2}\right)\tan\theta_i \\ \cos\angle DOF = \frac{\overline{OD}}{\overline{OF}} = \frac{\sqrt{R_o^2 - L^2}}{\sqrt{R_o^2 - L^2 + d_i^2}} \\ \sin\angle DOF = \frac{\overline{DF}}{\overline{OF}} = \frac{d_i}{\sqrt{R_o^2 - L^2 + d_i^2}} \\ \cos\angle BOC = \frac{\overline{BC}}{\overline{OB}} = \frac{\sqrt{R_o^2 - d_i^2}}{R_o} \\ \sin\angle BOC = \frac{\overline{OC}}{\overline{OB}} = \frac{d_i}{R_o} \\ \cos\left(\angle DOF - \angle BOC\right) = \cos\angle DOF\cos\angle BOC + \sin\angle DOF\sin\angle BOC \\ \cos\alpha = \cos\left(\angle DOF - \angle BOC\right) = \frac{\sqrt{R_o^2 - L^2}}{\sqrt{R_o^2 - L^2 + d_i^2}} \times \frac{\sqrt{R_o^2 - d_i^2}}{R_o} + \frac{d_i}{\sqrt{R_o^2 - L^2 + d_i^2}} \\ \times \frac{d_i}{R_o} \\ X_i = \overline{BF} \\ \overline{BF}^2 = \overline{OF}^2 + \overline{OB}^2 - 2 \times \overline{OF} \times \overline{OB} \times \cos\alpha \\ X_i^2 = \left(R_o^2 - L^2 + d_i^2\right) + R_o^2 - 2\sqrt{R_o^2 - L^2 + d_i^2}R_o\cos\alpha \\ \quad = 2R_o^2 - L^2 + d_i^2 - 2\left(\sqrt{R_o^2 - L^2} \times \sqrt{R_o^2 - d^2} + d_i^2\right) \end{cases} \tag{3}$$

In the formula: θ—Detects the central angle of the arc, °; α—The reference calculation angle, °; i—Ranging serial number; N—Total number of unilateral ranging sensors; θ_i—The original central angle of the unilateral i-th-ranging sensor, °; d_i—The horizontal and vertical distance between the detection point of the i-th ranging sensor and the center position (the $N + 1$) ranging sensor, mm; L—The length of the autonomous navigation platform is half, mm; R_o—An original detection radius, mm; X_i—The vertical distance of the detection point of the i-th-ranging sensor, mm.

$$\begin{cases} X_i = \sqrt{2R_o^2 - L^2 - d_i^2 - 2\sqrt{\left(R_o^2 - L^2\right)\left(R_o^2 - d_i^2\right)}} \\ = \left(R_o^2 - d_i^2\right) - \left(R_o^2 - L^2\right) \\ L_x = 2R_o\sin\frac{\theta}{2} \\ X_0 = \frac{1}{\tan\left(90 - \frac{\theta}{2}\right)}L_x \\ x_i = \begin{cases} x_i = \sqrt{\left(L_C + d + X_i\right)^2 + \left(d_i\right)^2} - R_m \ \ i0 \\ x_0 = L_C + d + X_0 - R_m \ \ i = 0 \end{cases} \\ a_i = 90 - \arctan\frac{L_C + d + X_i}{d_i} \end{cases} \tag{4}$$

In the formula: R_m—The original radius of the miniaturized circular arc array module, mm; d—The vertical distance between the original detected arc chord length and the side of the autonomous navigation platform, mm; L_C—Half width of the autonomous navigation platform, mm;x_0.Distance detected by sensors at center positions, mm; x_i. Detection distance of different position sensors.

Installation position under different vehicle width and test arc characteristics match

According to the modular design scheme of the arc array edge navigation sensor shown in **Figure 11**, the lateral installation position of the edge navigation sensor module is the only one, and the longitudinal center line of the N + 1 sensor at the center position is always installed coincidently with the horizontal center line of the autonomous navigation platform. However, the relative position between the detection arc and the side of the autonomous navigation platform, the width of the autonomous navigation platform, and the distance between the detection arc chord length and the side of the autonomous navigation platform determine the longitudinal installation position of the edge navigation sensor module, but the bottom straight line of the edge navigation sensor module should always coincide with or be installed in parallel with the longitudinal centerline of the autonomous navigation platform. As shown in **Figure 13**, the arc array edge navigation sensor module is made with 600 mm as the original vehicle width and 30 mm as the distance from the original detection arc chord length to the side of the autonomous navigation platform. In **Figure 13**, the longitudinal centerline of the n + 1 sensor at the center of the arc array edge navigation sensor module coincides with the horizontal centerline of the autonomous navigation platform. To cope with the installation position matching of autonomous navigation platforms with different widths and detection arcs with different features, a method for adjusting the longitudinal installation position of the edge navigation sensor module is established, as shown

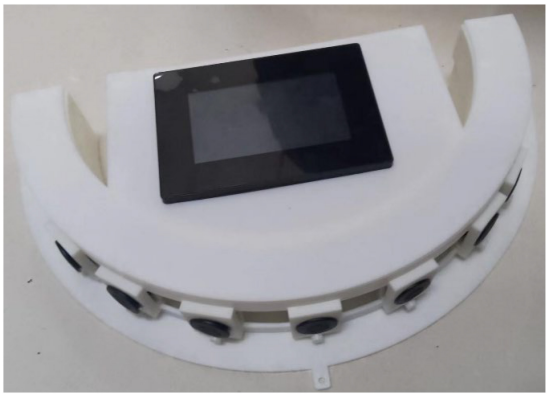


FIGURE 13
Physical drawing of a developed sensor.

in Equation (5). When $d_L < 0$, the arc array moves along the edge navigation sensor module toward the direction of detecting the center of the arc; when $d_L > 0$, the arc array moves along the edge navigation sensor module toward the detection arc direction.

$$d_L = (w_h + d_h) - (L_C + d) \begin{cases} d_L > 0 \text{ Adjust in the direction of arc} \\ d_L < 0 \text{ Adjust towards the center of the circle} \end{cases} \quad (5)$$

In the formula: d_L—distance to be adjusted, mm; w_h—actual width of the autonomous navigation platform, mm; d_h—actual chord length vertical distance from the side of the autonomous navigation platform, mm; d—Original chord length vertical distance from the side of the autonomous navigation platform, mm; L_C—Half the width of the original autonomous navigation platform, mm.

Adaptive calibration and the construction method of detecting an arc

It can be seen from **Figure 10** that the coupling adjustment layout method based on reducing the layout radius of the ranging sensor – adjusting the unequal center angle – increasing the detection distance is to realize the miniaturization design by adjusting the center angle and detection distance of the sensors at different positions. The angle of the circle center angle is quickly adjusted by sliding the angle-sliding adjusting plate equipped with the distance-measuring sensor at the position of the semicircular shell of the arc chute. Ju et al. (2017) used a switch type photoelectric switch to manually calibrate the detection distance, realizing the detection arc structure,

FIGURE 14
Physical drawing of a developed sensor.

resulting in cumbersome use, and the distance detection of the photoelectric switch is easily affected by environmental factors. In order to improve the convenience of using the arc array edge navigation sensor module, as shown in **Figure 14**, the PWM wave ultrasonic distance sensor (Dianyingpu DYP-A19-V1.0, Guangdong, China) is selected to detect the distance from the edge of the road. Based on the detection distance adjustment equation model of different position sensors, the trigger thresholds of different position sensors are automatically set to complete the automatic calibration of the detection points so as to realize the adaptive structure of the detection arc, as shown in **Figure 15**.

Error analysis and reduction of edge-relative pose detection

Error analysis of edge-relative pose detection

The relative pose detection performance of the arc array edge navigation sensor module directly affects the operation accuracy of the autonomous navigation platform. In order to improve the edge operation accuracy of the autonomous navigation platform, the edge relative pose detection error must be reduced. In view of this, it is necessary to detect the source of the relative pose error of the arc array edge navigation sensor module, and take corresponding measures to reduce the error impact. According to the edge navigation principle of the circular arc array and the coupling adjustment arrangement method of reducing the layout radius of ranging sensors – adjusting the center angle of non-equidistant circles – increasing the detection distance, it can be seen that the relative pose detection errors are mainly caused by two reasons. On the one hand, it is caused by the distance detection error of the ranging sensor and the uneven road edge plane; on the other hand, it is caused by the position and attitude solution model error of the arc array navigation along the edge, the too fast operation speed and the sudden change of the road edge plane position. It can be seen that the relative pose detection error along the edge is affected by multi-factor coupling. In order to further improve the

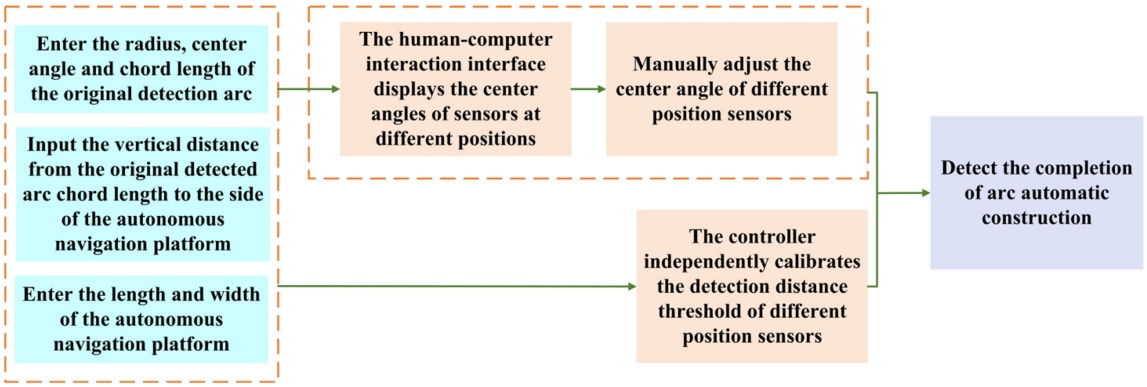


FIGURE 15
A schematic diagram of the construction method for detecting arc-adaptive calibration.

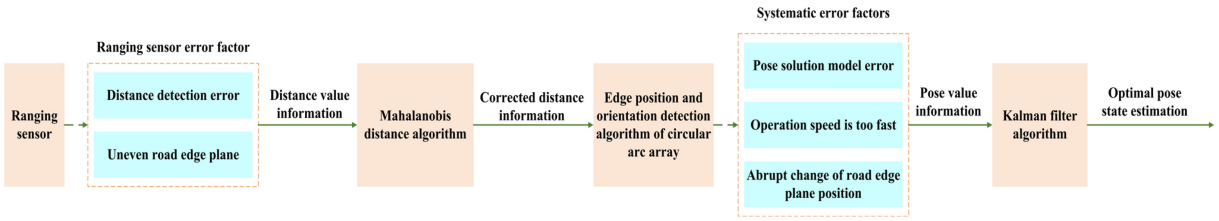


FIGURE 16
Multi-factor detection error compensation and a weakening scheme for an edge-relative pose.

smoothness and stability of the arc array edge navigation sensor module, as shown in Figure 16, a multi-factor error comprehensive compensation scheme is adopted. First, the Mahalanobis distance algorithm is used to weaken the influence of the distance detection error of the ranging sensor and improve the accuracy and stability of the determination of the relative edge position and attitude. Then, the Kalman filter algorithm is used to estimate the relative position and attitude along the edge to improve the detection accuracy and anti-interference performance.

Reduction of sensor-ranging error based on Mahalanobis distance algorithm

Because of the distance detection error and the uneven road edge plane, the ultrasonic distance sensor will have significant abnormal values in the distance measurement values, which will lead to the decline of the accuracy and stability of the determination of the relative position and attitude along the edge. In the process of ultrasonic sensor ranging, all ranging mean and covariance matrices are stable values calculated in the minimum covariance determinant estimation. Therefore, there is a significant difference between the Mahalanobis distance of abnormal values and the normal values in the calculated samples. Therefore, the Mahalanobis distance algorithm can be used to eliminate the interference of abnormal values of ultrasonic ranging sensors. In order to weaken the influence of environmental factors and the ultrasonic ranging sensor's own factors on ranging error, the arc array edge navigation sensor module uses Markov distance algorithm to detect and eliminate the abnormal ranging value of the ultrasonic ranging sensor. However, in the standard Mahalanobis distance algorithm, Mahalanobis distance is removed by calculating the mean sum and covariance matrix of the original ranging data. When the sample size of the distance measurement data is small and there are many outliers in the ranging, the standard Mahalanobis distance algorithm will make the calculated mean value and the covariance matrix of the original ranging data deviates from the outliers, resulting in the outliers being detected as normal values, which will lead to the incomplete elimination of the outliers. In view of this, the fast minimum covariance determinant algorithm is adopted to obtain a stable covariance matrix and a stable average vector, and then the stable Mahalanobis distance is calculated according to the standard Mahalanobis distance equation to eliminate the outliers of distance detection, as shown in Figure 17, in which we set the threshold of Mahalanobis distance approximated to the chi-square distribution, ranging numerical sample data as a constant. When the Mahalanobis distance of the calculated ranging value is greater than the set threshold, it will be regarded as an abnormal value and will be discarded immediately.

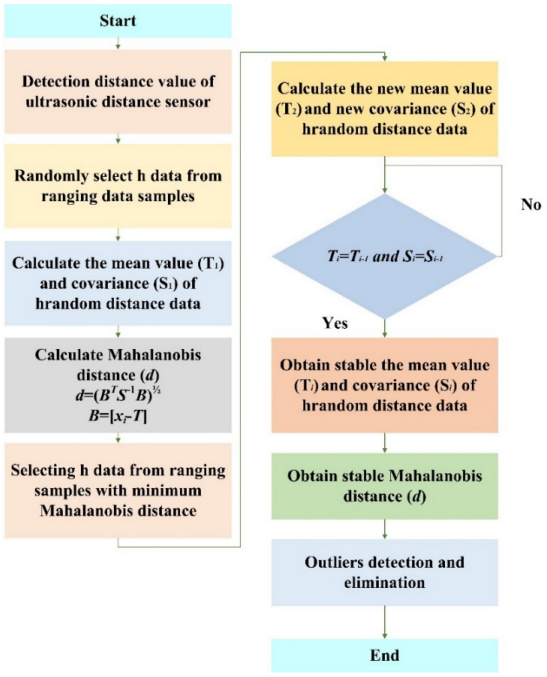


FIGURE 17
A flow chart of abnormal value Mahalanobis distance detection and the elimination method of the ultrasonic distance sensor: *d*, the Mahalanobis distance, h_i, the sample size, *T*, the mean value of the sample data, and *S*, the covariance matrix of the sample data.

Optimal estimation of a pose state based on the Kalman filter

The model error of pose solving for circular arc array edge navigation, too fast operating speed, and abrupt position along the road plane will further affect the position and attitude state detection accuracy. In order to ensure the operation accuracy of the autonomous navigation platform, the original position and attitude state detection information needs to be optimally estimated. Kalman filter algorithm is a recursive estimation algorithm in essence. It does not need to record observations and estimated historical data. It estimates the optimal value of the current state according to the estimated value of the last time state and the observed value of the current state in the system. It is widely used in the field of autonomous navigation and control. In view of this, the standard Kalman filter algorithm is used to smooth the relative position and attitude state information along the edge, improve the position and attitude detection accuracy and anti-interference performance, and weaken the influence of system error and environmental error, as shown in Figure 18. The specific steps are as follows: the first step is to detect the relative wayside pose data of the autonomous navigation platform; the second step is to establish the optimal estimation model of the relative heading deviation of the autonomous navigation platform and the relative lateral deviation of the center point of the autonomous navigation platform; Step 3: calculate the Kalman gain coefficient K_γ of the relative wayside heading deviation γ and the Kalman gain coefficient K_D of the lateral deviation of the relative path edge; the fourth step is to measure and predict the relative wayside pose of the autonomous navigation platform at time k (the current time) according to the original autonomous navigation platform's relative wayside pose at time k-1 (the previous time); the fifth step is to update the current time relative wayside lateral deviation estimation error and the current time relative wayside heading deviation estimation error for the next time relative wayside position and attitude information prediction of the autonomous navigation platform.

Integrated embedded module design

The arc array edge navigation sensor is designed to realize the real-time detection of the relative position and attitude of the autonomous navigation platform and the output of navigation control information through modular packaging. Therefore, on the basis of the miniaturization scheme of reducing the radius of the ranging sensor, adjusting the non-equal center angle and increasing the detection distance, it is necessary to design the hardware circuit and software program of the sensor system. In this section, we introduce the sensor module hardware system design and the sensor software program design, in turn.

Hardware design of a sensor module

In this paper, the arc array edge navigation sensor module is designed to accommodate up to 9 ultrasonic ranging sensors, as shown in Figure 19. In order to improve the anti-interference performance of environmental factors and realize the adaptive calibration structure of the detection arc, a PWM wave ultrasonic ranging sensor (Dianyingpu DYP-A19-V1.0, with the optimal accuracy of 1 mm, Guangdong, China) is selected to detect the distance from the road edge. The sensor uses the pulse width change time to measure the distance. It needs to use a microsecond timer to measure the pulse width time to complete the distance detection. At present, the commonly used PLC logic controller can only achieve millisecond timing. The number of timers of a 51 single-chip microcomputer controller is small and cannot support more than three ultrasonic distance sensors. However, there are more than eight timers in the STM32 MCU controller. In order to accommodate the distance measurement of nine ultrasonic ranging sensors at the same time, and increase the expansion performance of the arc array edge navigation sensor module. In this paper, stm32f407 embedded controller (punctual atomic core board, Guangdong, China) is selected to develop the hardware system of the sensor module. The

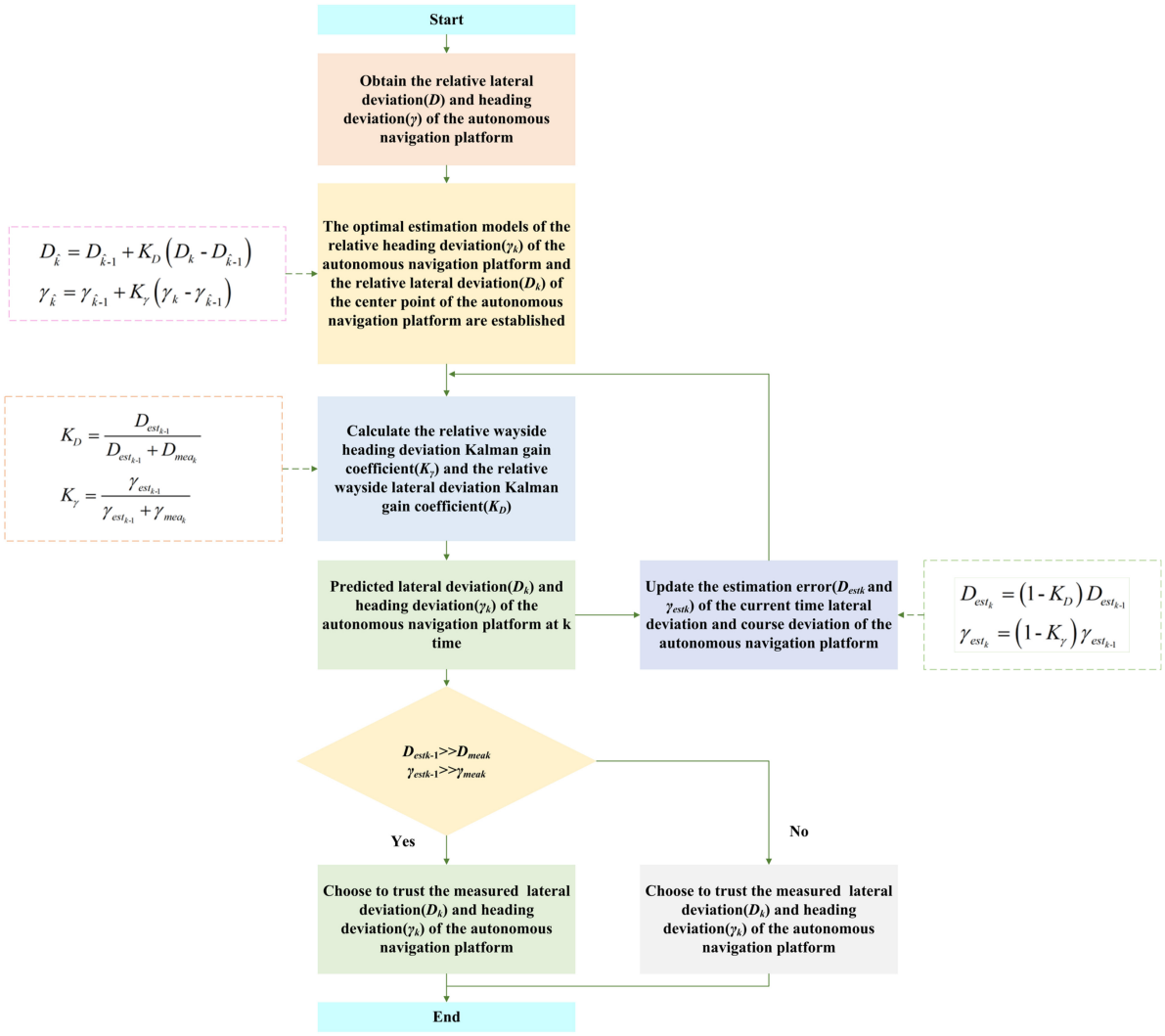


FIGURE 18
Kalman filter optimal estimation flow of a relative wayside pose: D_k. The observed lateral deviation of the current autonomous navigation platform relative to the road edge. $D_{\hat{k}}$, the estimated lateral deviation of the current autonomous navigation platform relative to the road edge. $D_{\hat{k}-1}$, estimated value of the lateral deviation of the autonomous navigation platform relative to the road edge at the last moment. K_X, Kalman gain coefficient of relative lateral deviation of the autonomous navigation platform. γ_k, the observed heading deviation of the current autonomous navigation platform relative to the road. $\gamma_{\hat{k}}$, estimated heading deviation of the current autonomous navigation platform relative to the road edge. $\gamma_{\hat{k}-1}$, estimated value of the heading deviation of the autonomous navigation platform relative to the road edge at the last time. K, Kalman gain coefficient of relative heading deviation of the autonomous navigation platform. D_{estk-1}, the estimation error of the lateral deviation of the autonomous navigation platform relative to the road edge at the last moment. D_{meak}, the current autonomous navigation platform-relative lateral deviation measurement error along the road. γ_{estk-1}, the estimation error of the heading deviation of the autonomous navigation platform relative to the road at the last moment. γ_{meak}, the current autonomous navigation platform relative along the road heading deviation measurement error. D_{estk}, estimation error of lateral deviation of the autonomous navigation platform relative to the road edge at the current time. γ_{estk}, the estimated error of the relative heading deviation of the autonomous navigation platform at the current time.

controller built in 14 timers. In order to improve the user-friendliness of the arc array edge navigation sensor module and realize the adaptive structure of detecting arcs, an HMI serial port display screen (4.3-inch, four-wire system, Guangdong, China) is added to the sensor module hardware system. The center angle, arc radius, and photoelectric switch number of the original detected arcs can be input by touch, as well as the length and width of the autonomous navigation platform, and complete the automatic calculation of the adjustment parameters of the center angle arrangement of different position sensors and the independent calibration of the distance between detection points. The output of relative path edge position and attitude state of the autonomous navigation platform is the key to the design of the arc array edge navigation sensor module. The TTL serial port circuit is set on the hardware system of the sensor module, which can realize the position and attitude state

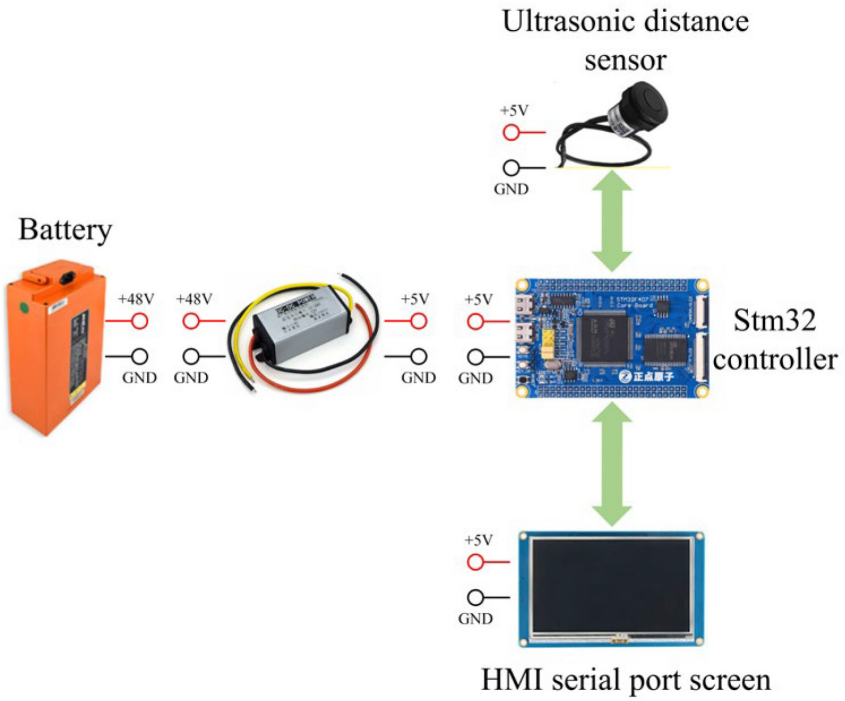


FIGURE 19
Hardware system composition of an arc array edge navigation sensor module.

TABLE 2 Manufacturing cost of an arc array edge navigation sensor module.

Type of sensors				
Arc array edge navigation sensor module			Lidar	Visual camera
Part name	Unit cost/$	Price range/$	Price range/$	Price range/$
Plastic shell	30	<100	>300	>400
STM32 controller	25			
HMI serial port screen	30			
Ultra sonic distance sensor	5			
Power conversion module	10			

and position and attitude deviation value in the form of RS232 or RS485 serial ports through different external modules so as to meet the control requirements of different lower computers. The cost of the arc array edge navigation sensor module designed in this paper is about US $100. As shown in Table 2, compared with the current laser radar and a visual camera, its price is relatively lower and its use is simpler.

Software design of the sensor module

In this paper, the arc array edge navigation sensor module completes the detection and output of the relative position and attitude of the autonomous navigation platform through integrated packaging. It is necessary to carry out a complete process packaging design for the software program of the sensor system, as shown in Figure 20. The sensor system software program follows the sequence of key parameter setting, manual adjustment of the center angle of sensors in different positions, synchronous ranging of multiple ultrasonic sensors, calculation of relative roadside position and attitude state and output of position and attitude state signal execution. Finally, the modular design of the autonomous edge navigation sensor is realized. The edge-navigation method based on the principle of the circular array can be controlled according to the lateral deviation and heading deviation of the autonomous navigation platform relative to the road edge, and, also, according to the nine position and attitude states of the autonomous navigation

10.3389/fpls.2022.892388

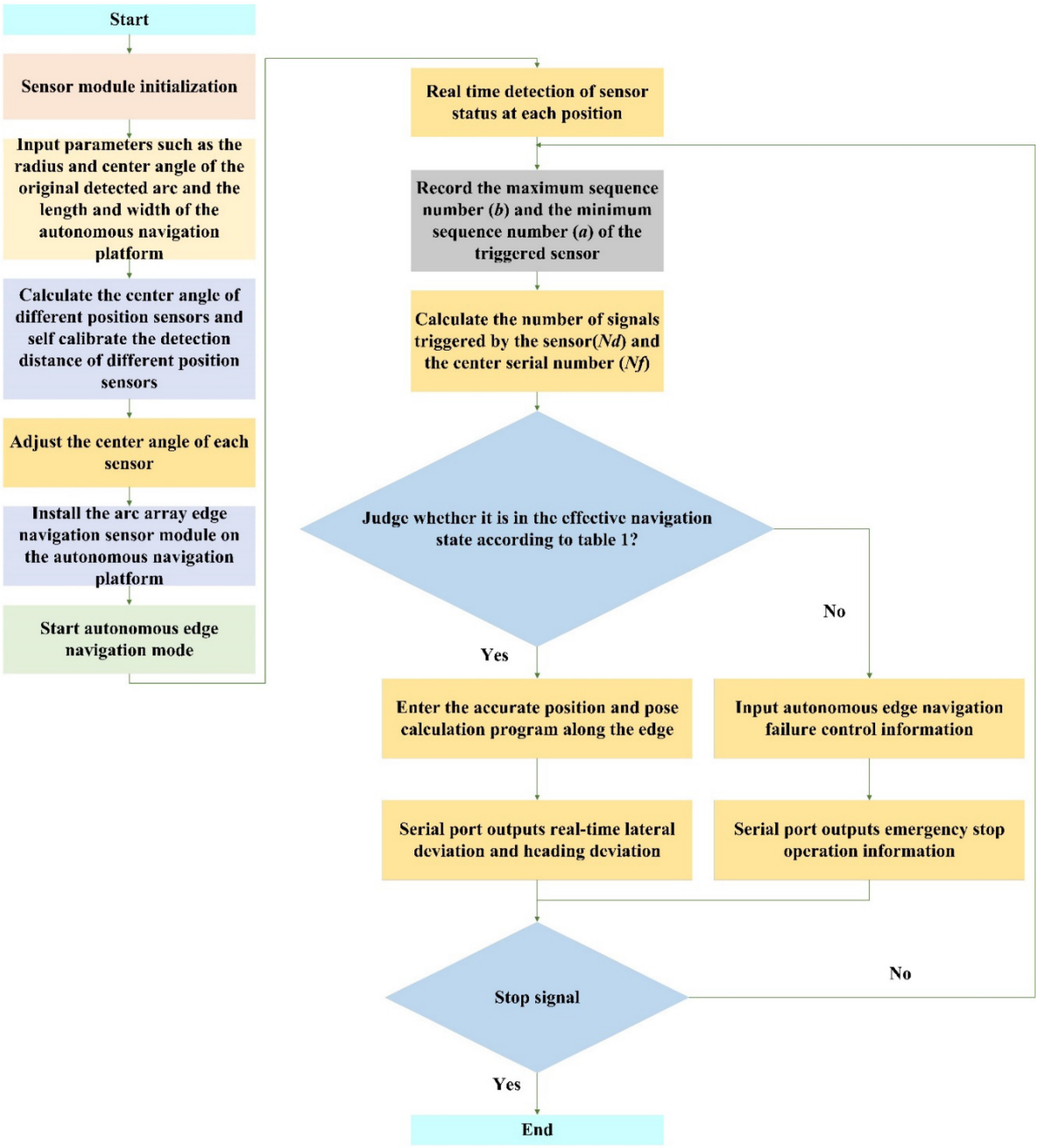


FIGURE 20
A software program flow of arc array edge navigation sensor module.

TABLE 3 A relative wayside pose information data format.

Name	Identifier	Lateral deviation of relative road edge			Course deviation of relative road edge			Pose type	Checksum
Bytes	0	1	2	3	4	5	6	7	8
Definition	0 × 01	sx	xx	yy	sx	xx	yy	xx	CRC

FIGURE 21
Performance testing.

platform relative to the road edge in Table 1. In order to ensure the safety of navigation along the border, the conditions not within the scope of Table 1 are considered as invalid conditions, which are used as the judgment of alarm signal output. In order to facilitate the user to detect the control signal output from the edge navigation sensor module of the arc array, the establishment data communication format shown in Table 3 is established. The data format uniformly adopts the hexadecimal compressed *BCD* code. The specific design is as follows: sxxxyy, sx is the symbol bit (sx = 00 indicates that the lateral deviation and heading deviation of the autonomous navigation platform relative to the edge of the road are positive, sx = 10 indicates that the lateral deviation and heading deviation of the autonomous navigation platform relative to the edge of the road are negative), xx represents integer digits, yy represents decimal digits, and the data content is the horizontal deviation and heading deviation of the autonomous navigation platform relative to the wayside. Where the 0th byte is 0 × 01 represents the frame identifier of the pose parameter message, the 1st, 2nd, and 3rd bytes represent the lateral deviation of the autonomous navigation platform relative to the curb, the 4th, 5th, and 6th bytes represent the heading deviation of the autonomous navigation platform relative to the curb, the 7th byte represents the status of the autonomous navigation platform relative to the curb, and the 8th byte is the checksum, which is used to verify the integrity and accuracy of the data.

Experiment

In order to verify the rationality and effectiveness of the arc array edge navigation sensor module, a detection arc (Ju et al., 2017) was built on the autonomous navigation platform with a length of 1,200 mm and a width of 600 mm to carry out different feature detection arc impact tests, environmental factor impact and comparison tests, bump obstacle impact and comparison tests, and dynamic detection performance and comparison tests so as to detect the error of heading deviation. The horizontal deviation detection error and variation coefficient are analyzed as evaluation indicators, as shown in Figure 21.

An arc influence test with different features

According to Equations (1) and (2) in this article, the position and orientation detection accuracy of autonomous curb line navigation is related to the setting parameters of the number of ultrasonic distance sensors (N_C), the detection arc radius (R), and the central angle (θ). In order to clarify the influence of key layout parameters on the accuracy of the position and orientation detection, furtherly, the number of photoelectric switches is selected as 5, 7, and 9, the detection arc radius is 3,300, 4,300, and 5,300 mm, and the central angle is 5°, 10°, and 15° as the test levels; under the setting conditions that the detection arc radius is 3,300 mm and the central angle is 5°, the number of photoelectric switches is 7 and the central angle is 5°, the number of photoelectric switches is 7 and the detection arc radius is 3,300 mm. The forward normal condition (lateral deviation D = 0 mm, heading deviation γ = 0°) in Table 1 is used to test the detection accuracy of the edge pose affected by a single factor in the stationary state of the autonomous navigation platform.

In order to meet different accuracy requirements of autonomous navigation along edges, and to facilitate users to choose a circular arc array layout that meets the actual requirements. The number of ultrasonic ranging sensors, the radius of circle arc and the central angle of circle are taken as test factors. Similarly, under normal working conditions, the number of ultrasonic ranging sensors is 5, 7 and 9. The radius of the detected arc is 330 mm, 4300 mm and 5300 mm, and the central angle is 5 degrees, 10 degrees and 15 degrees as the test level. The multi-factor coupling influences the accuracy of position detection along the edge.

A test on abrupt change of a concave convex along the road surface

Aiming at the problem that the photoelectric ranging sensor is easy to suffer from poor accuracy caused by environmental factors, such as light, reflecting surface color, and wall material, and to ensure that the detection distance of sensors at different positions can be independently calibrated to construct the detection arc requirements, the system selects the ultrasonic ranging sensor to develop the circular arc array edge navigation sensor. Because the ultrasonic distance sensor is a non-contact and wear-free detection of the detected object by using the acoustic medium, it has the characteristics of high frequency, short wavelength, and a

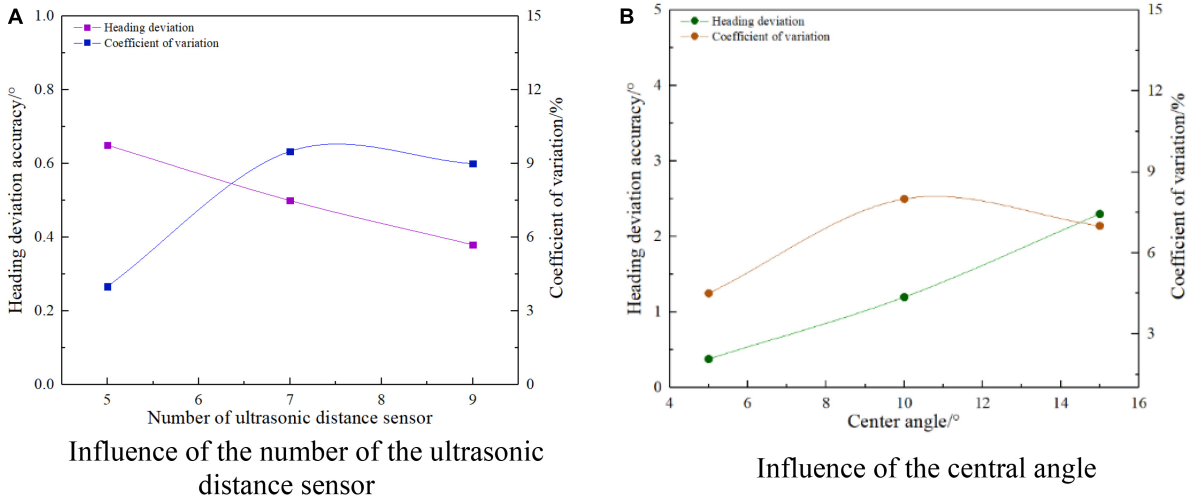


FIGURE 22
A single factor affects the accuracy of heading deviation detection.

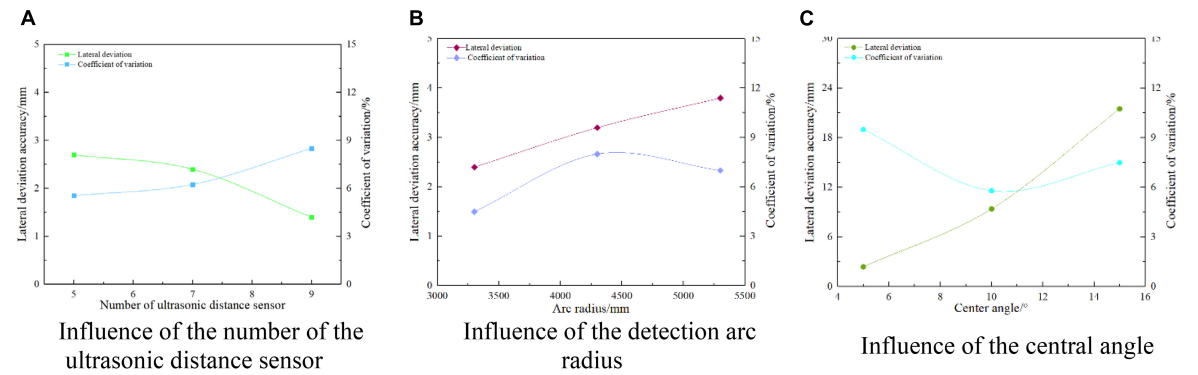


FIGURE 23
A single factor affects the accuracy of lateral deviation detection.

small diffraction phenomenon. It can detect transparent or colored objects, metallic or non-metallic objects, solid, liquid, powdery substances, especially objects through which light cannot pass; the detection performance is hardly affected by any environmental conditions. Therefore, environmental factors, such as illumination, reflector color, and wall material are not considered to affect the performance of the circular arc array edge navigation sensor. However, there will be bump changes in the actual road edge plane. At this time, the effective pose reflected by the arc array edge navigation sensor deviates from the actual situation. In order to verify the influence of bump mutation along the road plane on the detection effect of the autonomous navigation platform on the edge position and attitude state, a detection arc with radius R = 3,291 mm and the center angle θ = 20° and chord length L = 1,200 mm (Ju et al., 2017) was constructed with seven ultrasonic ranging sensors. The position and attitude detection performance of concave and convex changes along the road was tested respectively according to the positive normal condition in Table 1 as the initial position and attitude state.

Dynamic detection performance test

In order to verify the dynamic pose detection performance of the small integrated arc array edge navigation sensor module designed in this study, a detection arc with radius R = 3,291 mm and the center angle θ = 20° and chord length L = 1,200 mm (Ju et al., 2017) was constructed with seven ultrasonic ranging sensors. The test is carried out according to the independent position and attitude detection along the edge by walking in a straight line at the speed of 0.15, 0.25, and 0.35 m/s under the positive

TABLE 4 Multi-factor coupling affects the accuracy of heading deviation detection.

Serial number	Influence factor		Course deviationγ/°	Coefficient of variation/%
	Total number of ultrasonic distance sensor/pcs	Central angle /°		
1	5	5	0.63	4.0
2		10	1.25	5.5
3		15	1.90	4.5
4	7	5	0.40	9.5
5		10	0.80	8.0
6		15	1.25	9.0
7	9	5	0.32	9.0
8		10	0.63	9.5
9		15	0.94	9.5

position and normal condition in Table 1. According to the triangular geometric relationship between the detection arc and the autonomous navigation platform, it can be known that the distance from the center point of the autonomous navigation platform to the edge under the positive position and normal condition is 300 mm. In order to ensure the repeated measurement accuracy and eliminate the random error in the measurement process, repeat the test for five times, take the average value of the test results, and analyze the stability.

Results

A test of the influence of the arc on different features

It can be seen from Equations (1) and (2) that only the number of ultrasonic distance measuring sensors and the influence of the center angle are analyzed for the course deviation detection performance relative to the road edge. Figure 22 shows the influence of ultrasonic ranging sensor number and the center angle on the detection performance of heading deviation with a relative road edge. It can be seen from Figure 22A that the detection error of relative wayside heading deviation is not more than 1°, and the coefficient of variation is less than 10%. When the center angle is fixed, the detection error of relative wayside heading deviation decreases with the increase of the number of ultrasonic-ranging sensors, and the minimum detection error of relative wayside heading deviation is 0.4°. It can be seen from Figure 22B that the detection error of the relative wayside heading deviation is less than 3°, and the coefficient of variation is not greater than 8%. When the number of ultrasonic ranging sensors is fixed, the detection error of the relative wayside heading deviation increases with the increase of the center angle. When the center angle is 15°, the maximum detection error is 2.3°. To sum up, the number of ultrasonic ranging sensors is negatively correlated with the detection error of relative wayside heading deviation, and the center angle is positively correlated with the detection error of relative wayside heading deviation, which is consistent with the change of position and attitude detection deviation accuracy reflected in formulas (1) and (2). Therefore, under the condition of meeting the travel control time response, the number of ultrasonic ranging sensors should be increased and a small center angle should be set.

As shown in Figure 23, it is the result of the influence of the number of ultrasonic distance measuring sensors, arc radius, center angle, and the transverse deviation detection performance of the relative edge. It can be seen from Figure 23A that the transverse deviation detection error of the relative road edge is less than 3 mm, and the coefficient of variation is not greater than 9%. When the arc radius and the center angle are fixed, the transverse deviation detection error of the relative road edge decreases with the increase of the number of ultrasonic distance-measuring sensors, and the minimum transverse deviation detection error is 1.4 mm. It can be seen from Figure 23B that the detection error of the transverse deviation of the relative path is not more than 4 mm, and the coefficient of variation is less than 8%. When the number of ultrasonic distance-measuring sensors and the center angle is fixed, the detection error of the transverse deviation of the relative path edge increases, with the increase of the arc radius. When the arc radius is 5,300 mm, the maximum detection error has reached 3.8 mm. It can be seen from Figure 23C that the transverse deviation detection error of the relative road edge is not more than 30 mm, and the coefficient of variation is less than 10%. When the number of ultrasonic distance-measuring sensors and the arc radius is fixed, the transverse deviation detection error of the relative road edge increases, with the increase of the center angle. When the center angle is 15°, the maximum detection error is 21.5 mm. To sum up, on the premise of meeting the response accuracy requirements of autonomous navigation walking control, the ultrasonic ranging sensor should be appropriately increased, and the smaller arc radius and the center angle should be selected.

The performance of relative wayside heading deviation detection is related to the number of ultrasonic distance sensors and the center angle, while the performance of relative wayside lateral deviation detection is related to the number of ultrasonic distance sensors, arc radius, and center angle. Therefore, firstly, the influence of the number of ultrasonic distance-measuring sensors, the center angle, and the relative

10.3389/fpls.2022.892388

TABLE 5 Multi-factor coupling affects the accuracy of lateral deviation detection.

Serial number	Influence factor			Lateral deviation γ/mm	Coefficient of variation/%
	Total number of ultrasonic distance sensor/pcs	Central angle/°	Arc radius/mm		
1	5	5	3300	2.77	5.50
2			4300	3.61	4.35
3			5300	4.45	5.00
4		10	3300	10.86	6.50
5			4300	14.15	4.85
6			5300	17.44	7.00
7		15	3300	24.32	6.00
8			4300	31.70	4.85
9			5300	39.07	5.00
10	7	5	3300	2.41	7.40
11			4300	3.14	8.00
12			5300	3.87	7.00
13		10	3300	9.47	7.50
14			4300	12.34	6.95
15			5300	15.21	7.25
16		15	3300	21.22	8.35
17			4300	27.65	9.00
18			5300	39.07	8.50
19	9	5	3300	1.41	6.35
20			4300	1.84	6.00
21			5300	2.26	5.95
22		10	3300	5.53	6.50
23			4300	7.20	7.00
24			5300	8.88	7.50
25		15	3300	12.40	6.25
26			4300	16.13	6.00
27			5300	34.10	6.00

wayside heading deviation detection performance is analyzed, and then the influence of the number of ultrasonic distance-measuring sensors, the arc radius, the center angle, and the relative wayside lateral deviation detection performance is analyzed. As shown in Table 4, in order to further analyze the effect of the number of ultrasonic ranging sensors and the center angle coupling on the detection performance of the relative wayside heading deviation, SPSS software is used to fit the performance equation of relative wayside deviation detection affected by two factors. As shown in Equation (6), R^2 after linear fitting of relative wayside deviation detection error and wayside deviation variation coefficient is 0.905 and 0.708, respectively. The fitting of relative wayside deviation variation coefficient is not significant, but the regression of relative wayside deviation detection error fitting model is significant, which can better describe the test results.

Further analyze the effect of the number of ultrasonic distance-measuring sensors, arc radius, and center angle coupling on the detection performance of the lateral deviation of the relative road edge, as shown in Table 5. SPSS software is used to fit the performance equation of three factors affecting the detection of the lateral deviation of the relative road edge, as shown in Equation (7). After linear fitting, R^2 of the detection error of the lateral deviation of the relative road edge and the variation coefficient of the course deviation are 0.935 and 0.373, respectively. The regression of the fitting model of the measurement error of the lateral deviation of the relative road edge is significant, indicating that it is consistent with the actual situation and can better describe the test results, but the fitting of the variation coefficient of the lateral deviation of the relative road edge is not significant, The reason for this result is that the lateral deviation detected by the arc

10.3389/fpls.2022.892388

TABLE 6 Test results of detection performance of a normal road edge plane turning to a concave road edge plane.

Serial number	Normal road edge plane				Normal road edge plane turning to concave road edge plane					
	Accuracy of lateral deviation		Accuracy of heading deviation		Accuracy of lateral deviation		Accuracy of heading deviation		Time of duration	
	Error mean/mm	Coefficient of variation /%	Error mean/°	Coefficient of variation /%	Error mean/mm	Coefficient of variation /%	Error mean/°	Coefficient of variation /%	Time of mean/s	Coefficient of variation /%
1	30	6.50	3.5	9.46	40	9.30	4.5	8.50	40	7.90
2	28		2.8		45		4.0		45	
3	28		2.8		40		5.0		40	
4	30		2.5		50		4.0		35	
5	25		3.0		40		4.5		40	
Mean	28.20	/	3.03	/	43	/	4.4	/	40	/

TABLE 7 Test results of detection performance of a normal road edge plane turning to a convex road edge plane.

Serial number	Normal road edge plane				Normal road edge plane turning to convex road edge plane					
	Accuracy of lateral deviation		Accuracy of heading deviation		Accuracy of lateral deviation		Accuracy of heading deviation		Time of duration	
	Error mean/mm	Coefficient of variation /%	Error mean/°	Coefficient of variation /%	Error mean/mm	Coefficient of variation /%	Error mean/°	Coefficient of variation /%	Time of mean/s	Coefficient of variation /%
1	30	7.63	3.5	9.08	45	4.41	4.0	8.53	45	8.45
2	28		3.5		40		5.0		38	
3	28		3		42		4.3		35	
4	30		2.8		45		4.5		38	
5	25		3		43		4.0		40	
Mean	31	/	3.16	/	43	/	4.36	/	39.2	/

TABLE 8 Test results of detection performance of a concave road edge plane turning to a normal road edge plane.

Serial number	Concave road edge plane turning to normal road edge plane					
	Accuracy of lateral deviation		Accuracy of heading deviation		Time of recovery	
	Error mean/mm	Coefficient of variation/%	Error mean/°	Coefficient of variation/%	Time of mean/s	Coefficient of variation/%
1	35	8.48	3.0	8.13	3.0	8.13
2	28		2.8		2.5	
3	28		2.5		3.0	
4	30		3.0		2.5	
5	30		2.5		2.8	
Mean	30.20	/	2.76	/	2.76	/

TABLE 9 Test results of detection performance of a convex road edge plane turning to a normal road edge plane.

Serial number	Convex road edge plane turning to normal road edge plane					
	Accuracy of lateral deviation		Accuracy of heading deviation		Time of recovery	
	Error mean/mm	Coefficient of variation/%	Error mean/°	Coefficient of variation/%	Time of mean/s	Coefficient of variation/%
1	35	9.08	3.5	8.48	2.5	6.50
2	30		3.0		3.0	
3	35		2.8		2.8	
4	28		3.0		3.0	
5	30		2.8		2.8	
Mean	31.6	/	3.02	/	2.82	/

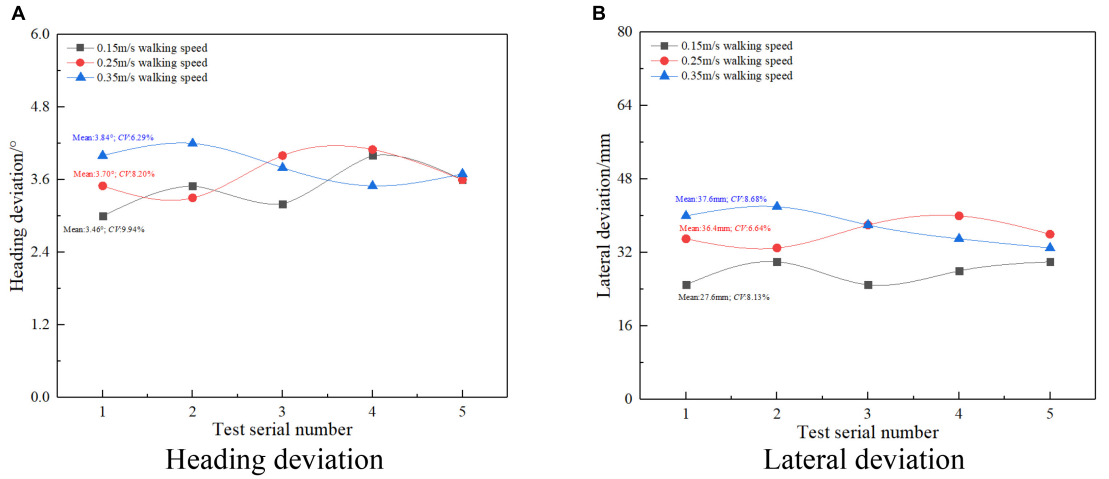


FIGURE 24
Dynamic test results.

array edge navigation method is a fluctuation range, but it will not affect the actual navigation effect. To sum up, we can predict the autonomous edge navigation detection accuracy of different feature detection arcs according to Equations (6) and (7). Users can meet the needs of different autonomous edge navigation accuracy by arranging different feature detection arcs.

$$\begin{aligned} \Delta\gamma &= -158 \times 10^{-3} N_t + 91 \times 10^{-3}\theta + 1.091 \\ R^2 &= 0.905 \\ \Delta\gamma_{CV} &= 1167 \times 10^{-3} N_t + 17 \times 10^{-3}\theta - 0.722 \\ R^2 &= 0.708 \end{aligned} \tag{6}$$

In the formula: $\Delta\gamma$— relative along the road heading deviation detection error, °; N_t—The number of ultrasonic ranging sensors; θ—Central Angle, °; $\Delta\gamma_{CV}$—The variation coefficient of relative roadside heading deviation detection error, %.

$$\begin{aligned} \Delta D &= -1628 \times 10^{-3} N_t + 2443 \times 10^{-3}\theta + 4 \times 10^{-2} R - 16.907 \\ R^2 &= 0.935 \\ \Delta D_{CV} &= 236 \times 10^{-3} N_t + 49 \times 10^{-3}\theta - 6.339 \times 10^{-5} R - 4.02 \\ R^2 &= 0.373 \end{aligned} \tag{7}$$

In the formula: ΔD—Lateral deviation detection error, mm; R—Radius of an arc, mm; ΔD_{CV}— Lateral deviation detection error variation coefficient, %.

A test on abrupt change of a concave convex along the road surface

Tables 6, 7 show the testing performance test of a normal road edge plane turning to a concave or a convex road edge plane. It can be seen from Table 6 that, when the normal road

edge plane turns to the concave road edge plane, the transverse deviation detection error relative to the road edge increases by about 10 mm and stabilizes at about 45 mm, and the heading deviation detection error relative to the road edge increases by 1° and stabilizes at about 5°. It can be seen from **Table 7** that, when the normal road edge plane turns to the convex road edge plane, the transverse deviation detection error of the relative road edge increases by about 8 mm, which is stable at about 40 mm, and the heading deviation detection error of the relative road edge increases by 0.5°, which is stable at about 4.5°. It can be seen from **Tables 6**, **7** that, when the normal path edge plane turns to the concave or convex path edge plane, the relative path edge attitude can maintain high-precision stable output for more than 30 s.

Tables 8, **9** show the testing performance test of a concave or convex road edge plane turning to a normal road edge plane. It can be seen from **Table 8** that, when the concave changed road turns to the normal road edge plane, the transverse deviation detection error of the relative road edge recovers to about 35 mm, and the heading deviation detection error of the relative road edge recovers to about 3.5°. It can be seen from **Table 9** that, when the convex changed road turns to the normal road edge plane, the transverse deviation detection error of the relative road edge recovers to about 30 mm, and the heading deviation detection error of the relative road edge recovers to about 3°. From **Tables 8**, **9**, it can be seen that, when the concave or convex road changes from the plane to the normal road edge plane, the relative road edge attitude can be restored to the detection state of the normal road edge plane within 5 s. To sum up, the arc array edge navigation sensor module can stably output high-precision position and attitude data for more than 30 s after turning from the normal road edge plane to the concave or convex road edge plane, which ensures the safety of autonomous edge navigation, effectively reduces the probability of collision accidents, and has certain fault tolerance and good anti-interference performance. The arc array edge navigation sensor module can quickly restore the normal position and attitude detection state after changing from the concave or convex path edge plane to the normal path edge plane, which has good real-time performance.

A dynamic detection performance test

Figure 24 shows the dynamic detection performance results of the small integrated arc array edge navigation sensor. The detection error of relative course deviation along the road is less than 4.5°, the coefficient of variation is less than 10%, the maximum coefficient of variation is 9.94%, and the minimum coefficient of variation is 5.59%. The detection error of lateral deviation of a relative road edge is less than 40 mm, the coefficient of variation is not more than 9%, the maximum coefficient of variation is 8.67%, and the minimum coefficient of variation is 6.64%. When the moving speed of the autonomous navigation platform increases from 0.15 to 0.35 m/s, the average detection error of the relative wayside heading deviation increases by 0.38°, and the average detection error of the relative wayside lateral deviation increases by 10 mm. As the walking speed of the autonomous navigation platform increases, the detection error of the relative wayside lateral deviation heading deviation increases slightly, but the variation coefficient of the relative wayside lateral deviation heading deviation is less than 10%. The performance of pose state detection is stable and real-time, which is not easily affected by the walking speed change of the autonomous navigation platform.

In addition, compared with Jia et al. (2015), to realize the edge detection method of a flat greenhouse road based on two-dimensional laser radar, the small integrated arc array edge navigation sensor is only affected by the unevenness and position mutation of the road edge plane, and has good adaptability to the gentle slope and uneven road surface on both sides of the road. Compared with the method of using machine vision to realize autonomous navigation path extraction (Wang et al., 2012; Hiremath et al., 2014; Malavazi et al., 2018; Yang et al., 2018; Inoue et al., 2019; Xu et al., 2021), the small integrated arc array edge navigation sensor uses ultrasonic distance sensor to detect the edge on the basis of the relative edge pose detection method based on the principle of the ideal target zone of the lateral center arc array. It avoids the influence of environmental factors, such as light, color of a reflecting surface, and wall material, and has good anti-interference ability. At present, the edge navigation method of a single ultrasonic or infrared photoelectric sensor (Feng et al., 2012; Zhao et al., 2012; Zhou, 2014; Wang and Liang, 2015) and linear arrangement of multiple ranging sensors to form an array module (Du, 2010; Xu et al., 2010; Yuan and Li, 2013) cannot realize the accurate position and attitude feedback of the autonomous navigation platform, and it is difficult to ensure the smoothness and control accuracy of the edge navigation. However, a small integrated arc array edge navigation sensor based on the principle of ideal target zone of lateral central arc array – the construction of the relative edge pose detection method can realize accurate pose state detection, and can improve the smoothness and stability of an autonomous navigation platform.

Conclusion

Based on the ideal target band principle of the lateral central arc array, this paper studies the relative edge pose detection method, and constructs a universal relative edge pose detection model with unknown number of sensors. Aiming at the sore point that the current arc array is too large to be integrated, a small integrated arc array navigation sensor module is designed by reducing the layout radius of the ranging sensor, adjusting the

center angle of the non-equal circle and increasing the detection distance. The main conclusions are as follows:

(1) This research has developed a small integrated arc array navigation sensor module with a cost of about US $100, which can accommodate at least nine ultrasonic ranging sensor groups, and proposed an autonomous construction method for detecting arcs with different characteristic parameters based on adaptive calibration of detection distance so as to improve the convenience and friendliness of users, and, at the same time, it can meet different requirements for autonomous edge navigation. It can also meet the requirements of low-cost, high-precision, and fast border navigation in greenhouse, animal, and plant factories and other environments.

(2) The experimental results of arc detection with different features show that the accuracy of the edge position and orientation navigation method based on the arc array is related to the key layout parameters of the detected arc. With the increase of the number of ultrasonic distance-measuring sensors, the detection errors of the heading deviation and the lateral deviation of the relative road edge are reduced. The detection accuracy of the heading deviation of the relative road edge is increased to 0.4°, and the detection accuracy of the lateral deviation of the relative road edge is increased to 3 mm. With the reduction of the center angle, both the heading deviation and the lateral deviation detection error of the relative curb decrease. When the center angle is 5°, the heading deviation detection accuracy of the relative curb reaches 0.38° and the lateral deviation detection accuracy of the relative curb reaches 2.4 mm. With the reduction of the arc radius, when the arc radius is 3,300 mm, the lateral deviation detection accuracy of the relative curb reaches 2.4 mm, and the heading deviation detection accuracy of the relative curb is not affected. Therefore, when setting the detection arc, the number of ultrasonic distance-measuring sensors should be appropriately increased, and the smaller arc radius and the center angle should be selected. In addition, the number of ultrasonic distance sensors, arc radius, and center angle significantly affects the detection accuracy of lateral deviation and heading deviation relative to the curb. Through the linear fitting equation, the prediction regression equation of the number of ultrasonic distance-measuring sensors and the heading deviation detection accuracy of the center angle and the relative edge is obtained. The prediction regression equation of the number of ultrasonic distance-measuring sensors, the circular arc radius, and the horizontal deviation detection accuracy of the center angle and the relative edge is obtained. The R^2 factor of the linear fitting is 0.905 and 0.935, respectively, which has a high fitting reality; this equation can be used to predict the detection accuracy of the detection arc along the edge with different layout feature parameters so as to quickly select the layout scheme suitable for the actual needs.

(3) The experiment on abrupt change of a bump on the road edge plane: when facing the operation environment of concave change and convex change on the road edge plane, when turning from the normal road edge plane to the concave change road edge plane or convex change edge plane, the transverse deviation detection error relative to the road edge increases by 10 mm, which is stable at about 45 mm; the heading deviation detection error relative to the road edge increases by 1°, which is stable at about 5°; and the relative road edge attitude can maintain high-precision and stable output for 30 s. The arc array edge navigation sensor module has certain fault tolerance.

(4) The dynamic detection performance test results show that, when the arc radius is 3,291 mm and the center angle of the circle is 20°, and the traveling speed of the autonomous navigation platform is 0.15 to 0.35 m/s, the detection errors of the lateral deviation and the heading deviation relative to the road edge are less than 40 mm and 4.5°, respectively. As the traveling speed of the autonomous navigation platform increases, the average detection error of the relative road edge heading deviation increases by 0.38°, and the average detection error of the relative road edge lateral deviation increases by 10 mm. However, the coefficient of variation is less than 10%. The dynamic position and attitude detection performance of the arc array edge navigation sensor module is relatively stable and has good real-time performance. It is less affected by the walking speed change of the autonomous navigation platform, and can be used for autonomous edge navigation control.

The small-scale integrated photoelectric arc array edge navigation sensor studied in this paper uses the ultrasonic distance sensor to establish the detection arc. Because the transmitted signal of the ultrasonic distance sensor is a divergent conical detection surface, the reflected signals of the adjacent ultrasonic distance sensors will interfere with each other, resulting in large errors in the detection process. The plane position of the ultrasonic distance sensor can be reasonably arranged to solve the problem of mutual interference between adjacent ultrasonic distance sensors. In addition, the ultrasonic ranging sensor will be affected by temperature, humidity, air pressure, air flow, and other factors, which will reduce the ranging accuracy and cause large errors in the position and attitude of the relative roadside. Subsequently, the influence of the above environmental factors can be weakened through the temperature and humidity compensation algorithm to further improve the adaptability of multi agricultural scenarios.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the first author (BX; 2668804078@qq.com).

Author contributions

BX, JL, and LC designed the study, conducted the trials, and analyzed the data. BX wrote the manuscript. YL provided test prototypes. HJ and LL polished the English grammar of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key R&D Program Project (2018YFD0700705), the Hunan Provincial Innovation and Entrepreneurship Program Investment Special (2019GK5059), the science and technology innovation Program of Hunan Province (2021RC5025), the National Natural Science Foundation of China (31971795), the Jiangsu Province Advantage Discipline Project (PAPD-2018-87), the Agriculture of Jiangsu University Supported by the Ministry of Equipment Engineering (4111680002), and the Jiangsu Provincial Key Laboratory of Agricultural Equipment and Intelligent High Technology (MAET202102).

Acknowledgments

We thank the Experimental Center of the School of Agricultural Engineering, Jiangsu University for providing the test site, and sincerely thank the editor and reviewers of the journal.

Conflict of interest

YL was employed by Intelligent Equipment Company of Hunan Xiangyuan Golden Spike.

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

References

Du, L. (2010). *Design and Research on Greenhouse Spray Robot Chassis*. Beijing: Beijing Forestry University.

Feng, Z., Xiao, S., Feng, L., and Gong, Y.-L. (2012). Research on wall-following navigation strategy and algorithm for mobile mechanical mouse. *Comput. Eng.* 38, 172–174.

Ferentinos, K. P., Katsoulas, N., Tzounis, A., Bartzanas, T., and Kittas, C. (2017). Wireless sensor networks for greenhouse climate and plant condition assessment. *Biosyst. Eng.* 153, 70–81. doi: 10.1016/j.biosystemseng.2016.11.005

Gao, G., and Ming, L. (2014). Navigating path recognition for greenhouse mobile robot based on K-means algorithm. *Trans. Chin. Soc. Agric. Eng.* 30, 25–33. doi: 10.1109/ICNSC.2014.6819668

Han, J., Park, C., Kwon, J. H., Lee, J., Kim, T. S., and Jang, Y. Y. (2020). Performance evaluation of autonomous driving control algorithm for a crawler-type agricultural vehicle based on low-cost multi-sensor fusion positioning. *Appl. Sci.* 10:4667. doi: 10.3390/app10134667

Hiremath, S. A., van der Heijden Gerie, 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

Inoue, K., Kaizu, Y., Igarashi, S., and Imou, K. (2019). The development of autonomous navigation and obstacle avoidance for a robotic mower using machine vision technique. *IFAC PapersOnLine* 52, 173–177. doi: 10.1016/j.ifacol.2019.12.517

Jia, S., Li, J., Qiu, Q., and Tang, H. (2015). New corridor edge detection and navigation for greenhouse mobile robots based on laser scanner. *Trans. Chin. Soc. Agric. Eng.* 31, 39–45.

Ju, J. (2017). *Designofcurb-following greenhouse mobile plat formbase donarcarray of photo electricswithes*. Zhenjiang: JiangsuUniversity.

Ju, J., Liu, J., Li, N., and Li, P. (2017). Curb-following detection and navigation of greenhouse vehicle based on arc array of photoelectric switches. *Trans. Chin. Soc. Agric. Eng.* 33, 180–187.

Li, Y., Xu, J., Wang, M., Liu, D., Sun, H., and Wang, X. (2019). Development of autonomous driving transfer trolley on field roads and its visual navigation system for hilly areas. *Trans. Chin. Soc. Agric. Eng.* 35, 52–61. doi: 10.1155/2019/6738594

Liang, M. H., He, Y. F., Chen, L. J., and Du, S. F. (2018). Greenhouse environment dynamic monitoring system based on WIFI. *IFAC PapersOnLine* 51, 736–740. doi: 10.1016/j.ifacol.2018.08.108

Malavazi, F. B. P., Guyonneau, R., Fasquel, J. B., Lagrange, S., and Mercier, F. (2018). Lidar-only based navigation algorithm for an autonomous agricultural robot. *Comput. Electron. Agric.* 154, 71–79. doi: 10.1016/j.compag.2018.08.034

Soiket, M., Oni, A. O., and Kumar, A. (2019). The development of a process simulation model for energy consumption and greenhouse gas emissions of a vapor solvent-based oil sands extraction and recovery process. *Energy* 173, 799–808. doi: 10.1016/j.energy.2019.02.109

Takai, R., Yang, L., and Noguchi, N. (2014). Development of a crawler-type robot tractor using rtk-gps and imu. *Eng. Agric. Environ. Food* 7, 143–147. doi: 10.1016/j.eaef.2014.08.004

Wang, H., and Liang, W. (2015). Design of intelligent cleaner based on MSP430. *J. Southwest Univ. Natl.* 41, 593–598.

Wang, X., Han, X., Mao, H., and Liu, F. (2012). Navigation line detection of tomato ridges in greenhouse based on least square method. *Trans. Chin. Soc. Agric. Mach.* 43, 161–166.

Xu, H., Li, S., Ji, Y., Cao, R., and Zhang, M. (2021). Dynamic obstacle detection based on panoramic vision in the moving state of agricultural machineries. *Comput. Electron. Agric.* 184:106104. doi: 10.1016/j.compag.2021.106104

Xu, M., Chai, Z., and Xu, W. (2010). Wall-following control of a mobile robot with fuzzy Q-learning. *Electr. Mach. Control* 14, 83–88.

Yang, S., Mei, S., and Zhang, Y. (2018). Detection of maize navigation centerline based on machine vision. *IFAC PapersOnLine* 51, 570–575. doi: 10.1016/j.ifacol.2018.08.140

Yuan, Y., and Li, J. (2013). Precision control of mobile robot in wall-following behavior based on fuzzy logic. *World Sci Tech R D* 704–708.

Yunpeng, J., Gang, L., and Zhikun, J. (2019). GNSS dual antennas combined with AHRS are used to measure farmland topography. *Trans. Chin. Soc. Agric. Eng.* 35, 166–174.

Zhao, T. C., Yu, T., and Huang, D.-H. (2012). The application of sensor technology in intelligent mobile (cleaning) robotic devices. *Dev. Innov. Mach. Electr. Prod.* 25, 7–9.

Zhou, H. (2014). Mobile Robot wall-following navigation based on a single ultrasonic sensor. *Tech. Acoust.* 33, 243–246.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 September 2022
DOI 10.3389/fpls.2022.881904

OPEN ACCESS

EDITED BY
Daobilige Su,
China Agricultural University, China

REVIEWED BY
Pedro Pablo Gallego,
University of Vigo, Spain
Abbas Atefi,
California Polytechnic State University, United States

*CORRESPONDENCE
Jizhan Liu
1000002048@ujs.edu.cn

SPECIALTY SECTION
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

RECEIVED 23 February 2022
ACCEPTED 25 July 2022
PUBLISHED 20 September 2022

CITATION
Jiang Y, Liu J, Wang J, Li W, Peng Y and Shan H (2022) Development of a dual-arm rapid grape-harvesting robot for horizontal trellis cultivation. *Front. Plant Sci.* 13:881904. doi: 10.3389/fpls.2022.881904

Development of a dual-arm rapid grape-harvesting robot for horizontal trellis cultivation

Yingxing Jiang, Jizhan Liu*, Jie Wang, Wuhao Li, Yun Peng and Haiyong Shan

Key Laboratory of Modern Agricultural Equipment and Technology, Jiangsu University, Zhenjiang, China

It is extremely necessary to achieve the rapid harvesting of table grapes planted with a standard trellis in the grape industry. The design and experimental analysis of a dual-arm high-speed grape-harvesting robot were carried out to address the limitations of low picking efficiency and high grape breakage rate of multijoint robotic arms. Based on the characteristics of the harvesting environment, such as the small gap between grape clusters, standard trellis, and vertical suspension of clusters, the configuration of the dual-arm harvesting robot is reasonably designed and analyzed, and the overall configuration of the machine and the installation position of key components are derived. Robotic arm and camera view analysis of the workspace harvesting process was performed using MATLAB, and it can be concluded that the structural design of this robot meets the grape harvesting requirements with a standard trellis. To improve the harvesting efficiency, some key high-speed harvesting technologies were adopted, such as the harvesting sequence decision based on the "sequential mirroring method" of grape cluster depth information, "one-eye and dual-arm" high-speed visual servo, dual arm action sequence decision, and optimization of the "visual end effector" large tolerance combination in a natural environment. The indoor accuracy experiment shows that when the degree of obscuration of grape clusters by leaves increases, the vision algorithm based on the geometric contours of grape clusters can still meet the demands of harvesting tasks. The motion positioning average errors of the left and right robotic arms were (*X*: 2.885 mm, *Y*: 3.972 mm, *Z*: 2.715 mm) and (*X*: 2.471 mm, *Y*: 3.289 mm, *Z*: 3.775 mm), respectively, and the average dual-arm harvesting time in one grape cluster was 8.45 s. The field performance test verifies that the average harvesting cycle of the robot with both arms reached 9 s/bunch, and the success rate of bunch identification and harvesting success rate reached 88 and 83%, respectively, which were significantly better than those of existing harvesting robots worldwide.

KEYWORDS

grape, standard trellis, sequential mirroring, depth threshold segmentation, one-eye and dual-hands visual servo

Introduction

Grapes are soft-skinned, juicy berries that occupy an important position in the world of fruit production. In 2020, the total area of vineyard cultivation worldwide was 7.3 million hectares, there were 50 million tons of grapes produced in the world, and China produced approximately 20 million tons, creating substantial economic value for people worldwide (Organisation internationale de la vigne et du vin [OIV], 2020; Isrigova et al., 2021). Due to growing labor shortages, the need for harvesting robots for fresh grapes has become increasingly urgent.

Trellis grapes are mainly used for fresh fruit consumption and are extremely difficult to harvest because of the need to ensure the integrity of grape clusters and soundness of grapes for transportation and marketing requirements (Ozkan et al., 2007; Jianying et al., 2014; Zhu et al., 2014). Traditional trellis grape harvesting operations rely mainly on manual work performed by two hands to finish working together, one hand to support and the other to cut grape stems, to complete one grape harvesting process (Piazzolla et al., 2016; Wang et al., 2017). This harvesting model is both inefficient and has high labor costs and will not meet the rapid harvesting standard of the future grape industry. Grape trellis configurations are mostly horizontal in Asia, the planting height is as high as 2 m, and the harvesting point of grape stems is usually 1.8 m above the ground. Traditional single-arm harvesting robots have deficiencies such as long harvesting cycles, poor moving flexibility, and inaccurate fruit harvesting accuracy, and they cannot meet the requirements of grape harvesting in standard trellises (Possingham, 2006; Suvoéarev et al., 2013; Williams and Fidelibus, 2016). Therefore, a highly efficient harvesting robot must be designed for standard trellis grapes to address the embarrassing gap of a lack of reliable harvesting machines in the grape-growing industry.

At present, researchers worldwide are still in the exploratory stage of research on harvesting machinery for grapes on trellises, and their research methods mainly revolve around visual positioning identification of grape clusters and the design of end-effector configurations (Luo et al., 2016; Liu et al., 2019; Tang et al., 2020; Kalampokas et al., 2021; Majeed et al., 2021; Peng et al., 2021). Facing the growth characteristics of different types of fruits and vegetables, researchers have developed multiple types of picking equipment. Mehta et al. (2014) proposed a cooperative vision servo controller for autonomous harvesting to adjust the position of the end effector according to the real-time position of fruit and, to a certain extent, to weaken the interference of the complex environment in the harvesting process. Levin and Degani (2019) proposed a modular design of an agricultural robot structure by examining the phenomena of low reusability and narrow applicability of the harvesting robot structure, which has resulted in a large improvement in harvesting time and fruit-harvesting success rate. Wang et al. (2019) proposed an optimization method of harvesting posture to address the randomness of the citrus growth direction on stalks and designed an occluding end effector with a success rate of fruit stalk-shearing up to 89% and a harvesting success rate of the best posture up to 74%. Kurtser and Edan (2020) used a TSP approach to plan a work sequence and path of sensing and harvesting tasks for a bell pepper-harvesting robot and concluded that planning a series of tasks can reduce costs by 12%. These equipments and methods were only commissioned in the laboratory and not in a realistic agricultural environment (Kurtser and Edan, 2020).

Compared to single-arm robots, harvesting robots that use a two-armed operational strategy are more advantageous in grape trellises. The dual-arm robot extends up to 2.5 m and can cover all grape-growing areas of a standard trellis, and the harvesting efficiency is much higher than that of traditional robots. Zhao et al. (2016b) designed and tested a dual-arm frame equipped with two 3 DoF (degree of freedom) manipulators and two different types of end effectors used to pick tomatoes and exchanged the operator's commands and displayed the state information of the robot. Ling et al. (2019) developed a dual-arm cooperative approach for a tomato-harvesting robot using a binocular vision sensor, and with vacuum cup grasping and wide-range cutting, the success rate of robotic harvesting reached 87.5%, while the harvesting cycle time was more than 30 s. Yu et al. (2021) used an autonomous humanoid robot for apple harvesting. It shows success rates of 82.5 and 72% for the apple recognition and harvesting functions, respectively; however, the apple-harvesting time is more than 30 s, and it has a rough structure and end effector. The authors concluded that although some progress has been made in the development of current grape-harvesting robots, further research is essential. Dual-arm harvesting robots can substantially improve operational efficiency, but there is still a lack of integrated harvesting robots in grape harvest production.

Grape clusters planted with trellises are mostly suspended on top of trellises, and the distance range from the cutting point of the fruit stalk to the top of the trellis is 30–100 mm, resulting in a small space for the upper limit activity of the robotic arm, which makes harvesting difficult and requires higher precision in identifying fruit clusters (Vrochidou et al., 2021). To accomplish efficient grape harvesting in standard trellis complex environments, our research group invented a dual-arm grape-harvesting robot for high standard trellis environments. Its harvesting structure used an RGB-D camera for the environmental field of view scanning and obtained the spatial information of grape-harvesting points and transmitted it to a dual robotic arm control system. The robot is a modular design. Facing different fruit and vegetable harvesting requirements, it only changes the structure of end effectors and adjusts the parameters of the vision recognition algorithm to quickly achieve a variety of fruit-harvesting tasks. Robotic harvesting operations through unmanned control have high harvesting

quality and harvesting efficiency. They can significantly reduce the labor burden in grape harvesting and improve the efficiency of grape-harvesting operations (Zhao et al., 2016a; Ling et al., 2019; Seol et al., 2020).

Therefore, a dual-arm rapid grape-harvesting robot for the horizontal trellis was designed and analyzed in this article. This robot is integrated with a variety of sensors and actuators to enable unmanned operation processes. In section "Parameters of the horizontal trellis environment," the horizontal trellis environment is introduced. In section "Overall structure of the dual-arm rapid harvesting robot," the hardware and software architecture design of the robot for rapid harvesting is described. In section "Key technologies of dual arm rapid grape-harvesting robot," we introduced the key technologies of the robot. First, we propose a "one eye-dual-arm" high-speed parallel harvesting strategy based on the structural parameters of the horizontal trellis. The position of the camera in relation to the two arms was also determined. Then end effectors and the vision algorithm were optimized for rapid recognition and harvesting process implementation. The combination of end effectors and a vision algorithm substantially improves the tolerance for errors. Finally, a dual-arm harvesting strategy based on depth values is proposed to achieve a harvesting sequence and the division of operation space by spatially symmetrical segmentation. For the area where the two arms are prone to collision, we established the danger area and safety area. In the danger area, the two arms will use an asynchronous master–slave dual-robotic arm anticollision harvesting strategy. In section "Experiments," we present indoor accuracy experiments and field performance experiments. In section "Conclusion," some conclusions are provided. Meanwhile, the existing work deficiencies and future research work are discussed.

Materials and methods

Parameters of the horizontal trellis environment

The viticulture mode in horizontal trellises is the grape tree-planting method, in which the bottom of the trellis is supported by pillars, and the top is pulled by cross bars or lead wires to form a net-like shelf surface, and branches and vines grow on the shelf (**Figure 1**). A horizontal trellis has the advantages of ventilation, light penetration, easy branch management and high production. It has become one of the main modes of fresh grape cultivation.

The horizontal trellis is divided into two upper and lower layers by pulling a wire mesh at the top. The upper layer allows vine branches to grow and spread, confining a large number of branches and leaves to the upper area, while grapes grow by gravity and hang vertically downward, achieving separation between fruit, branches, and leaves. After several measurements, the height of grape clusters to the top of the trellis is usually 20–120 mm, and the height of the bottom of grape clusters to the ground is 1,700–1,900 mm.

Fresh grapes, as ready-to-eat fruit, need to meet the integrity and aesthetics of the bunches for later sale and eating, so there are higher operational standards for harvesting fresh grapes. Grape-harvesting methods with traditional trellises rely on manual hand harvesting, with one hand supporting grape bunches and the other hand shearing the fruit stem, which is harmful to health because of the long hours spent harvesting with a head-up posture. Based on this horizontal trellis, there is an urgent need to design an intelligent grape-harvesting robot for standard trellises to replace manual labor to complete tedious tasks. Most traditional fruit- and vegetable-harvesting robots use a single mechanical arm as the harvesting servo mechanism, resulting in extremely low single-cycle harvesting efficiency that is much lower than the manual operation efficiency and cannot meet the requirements of the grape industry.

We completed a study of horizontal trellises for fresh grapes in different vineyards in Jurong City, Jiangsu Province, China (119.25852°E, 31.88404°N). Standard grape trellises have many unique structural characteristics that harvesting robots need to adapt. In this particular working environment, the harvesting robot is required to meet the following design requirements.

(1) Based on the horizontal trellis structure and the vertical growth of grapes, the harvesting width, depth value of recognition range, walking step length and another factors as key parameters of this robot. Hand-eye combination configuration and the harvesting posture determine the range of this robot and the end-effector.For these special requirements from the environment structure, the analysis of robot construct with multi-parameters fusion becomes the central issue.
(2) The position of the camera relative to the robotic arms was determined to ensure that all grape-harvesting targets were fully integrated into the field of view in the camera and robotic arm harvesting range.The combined relationship between the camera and dual arms becomes the key point.
(3) The distribution of grape growth was random in the standard trellis. The harvesting robot needs to quickly identify grape targets and assign harvesting tasks to the two robotic arms to accomplish rapid and accurate visual servoing. Harvesting task assignment is an important prerequisite for visual servo.
(4) The robot needs a reasonable harvesting strategy. It can respond to unstructured environments in real time. It is a crucial technology for the robot to make the harvesting motion smoothly, accurately and at high speed.

FIGURE 1
Vineyard in Jurong City.

Overall structure of the dual-arm rapid harvesting robot

Hardware structure

Single-arm harvesting robots have some shortcomings, including a small operating width (1–1.5 m) and low harvesting efficiency (average of 25 s/cycle). Most robots rely on tractors for towing or rail transport (Kootstra et al., 2021). They are unable to navigate autonomously in response to agricultural environment changes. Therefore, it is extremely important to develop a robot with high harvesting performance, multisensor integration, and real-time sensing of environmental changes.

Figure 2 shows the developed dual-arm rapid grape-harvesting robot. Its structure includes a RealSense D435i depth camera, two 6 DoF robotic arms, and mobile tracked chassis. The RealSense D435i depth camera is mounted on top of the robot. This ensures that the camera obtains as much of the field of view as soon as possible. The RealSense D435i depth camera acquires the spatial coordinates of grape clusters by shooting a standard trellis environment. The two robotic arms are distributed with the camera mounting position as the center of symmetry. To ensure that the dual-arm working space covers the grape-growing space within the standard trellis, two robotic arms are mounted on both sides of the robot. Many sensors are integrated into the control box (Jetson Nano, STM32, robotic arm controllers). The camera and two robotic arms are mounted *via* steel to a mobile tracked chassis. To acquire a point cloud of grape trees in a reasonable view, a SICK 2D radar system is mounted on the front of the tracked chassis. Multiple electrical systems are integrated into a robot, and this robot can handle various requirements in a nonstructural agricultural environment.

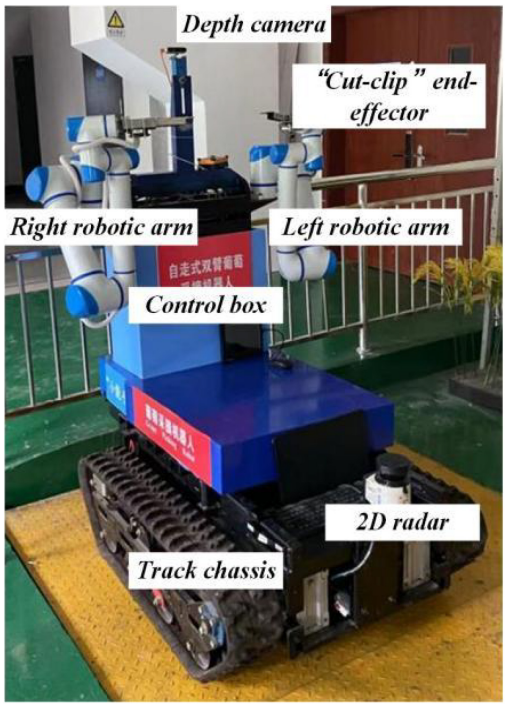


FIGURE 2
Dual-arm rapid grape-harvesting robot.

System architecture

A Nvida Jetson Nano developer kit is used as the center of the decision system. Its small size and powerful computing power meet the needs of running programs in harvesting (GPU: 128-core NVIDIA Maxwell, CPU: Quad-CoreARM Cortex-A57 MPCore). RealSense D435i depth camera is used as the main sensor to obtain environmental information (RGB images 1,920 × 1,080 in resolution, depth images 1,280 × 720 in resolution, with a FOV of 69° × 42°). It is manufactured by Intel, United States. It is able to cover a wider area and reduce more blind spots. The robotic arms use Techsoft TB6-R5 (Techsoft, Shenzhen, China). TB6-R5 has a payload of 5 kg and repeatable position accuracy of up to ±0.05 mm. Each robotic arm has its own

controllers, and controllers receive their respective harvesting tasks and control independently. Each arm is equipped with a cut-clip end effector. They hold grape clusters while cutting grape stems. The chassis is manufactured by Sangpu Agricultural Machinery Co., Changzhou, China. This robot uses a 2D radar (LMS-111, Sick, Germany) with TOF distance detection and enables accurate measurements in a complex field environment.

The dual-arm grape-harvesting robot consists of four main units: (1) a visual recognition system, (2) a decision system, (3) a servo harvesting system, and (4) a walking chassis system. The four units communicate with each other and work together to harvest the target fruit based on visual information.

To adapt to the field environment in agriculture, agricultural robotic systems are often required to have strong integration. **Figure 3** illustrates the details of the control system for the whole robot. In the hardware section, Jetson Nano is responsible for key aspects such as image processing, motion information transmission, and communication between each hardware unit. RealSense D435i is connected to Jetson Nano *via* USB and sends the 3D information acquired to Jetson Nano in real time. These images are segmented, and the algorithm extracts contour information within ROS (robot operating system). The robotic arm (Techsoft, TB6-R5, CHN) communicates with the controller in real time *via* an EtherCAT bus. After the robotic arm moves to the target coordinate, it sends a signal to Jetson Nano. Jetson Nano sends control commands to STM32 through serial ports. STM32 controls the opening and closing of the electric gripper. When there are no harvesting targets in the camera field of view, the chassis moves forward some distance. Until the camera requires harvesting targets again, the chassis stops moving, and then, the next harvesting cycle begins.

In the software part, ROS is currently the most popular control system in robots. It is able to manage and transmit multiple sensor data. The data of the camera, robotic arms, grippers, and chassis are defined as nodes. These nodes subscribe to each other through topics for data delivery. The overall software component allows for a rapid response to agricultural environmental changes (**Figure 4**).

Key technologies of dual arm rapid grape-harvesting robot

"One eye-dual hand" structure based on horizontal trellises

A hand-eye structure is the basis of robot vision servo control. At the same time, "eye-in-hand" usually requires a camera at the end of the arm. This results in a small camera field of view and cannot capture all the harvesting targets in the horizontal trellis. As shown in **Figures 5A–C**, three kinds of "eye-in-hand" structure occurs in different scenes.

Therefore, the special "one eye-dual hand" structure is proposed. This structure ideally ensures full coverage of all grape clusters in grape-growing space under a horizontal trellis. The rational arrangement of the mounting position relationship between the two arms and camera becomes the core of the robot harvesting structure. To obtain as many grape clusters as possible, the camera field of view needs to match the dual-arm working space (**Figure 6**; Barth et al., 2016; Seyyedhasani et al., 2020; Chen et al., 2021).

A robotic arm working space is usually defined as a spherical space to simplify the problem in traditional research. However, the 6 DoFs robotic arm consists of motors and links. It does not have an ideal spherical shape of a robotic arm because of the difference in length and orientation of links. It looks more like a rugby ball in a working space. Therefore, it would be more reasonable to analyze robotic arm working space using the ellipsoidal spherical space equation.

$$V = V_l + V_r - V_k \tag{1}$$

$$V_l = V_r = \iiint \left(\frac{x^2}{b^2} + \frac{y^2}{a^2} + \frac{z^2}{c^2}\right) dydxdz \tag{2}$$

As shown in **Figure 6A**, the left robotic arm working space $\boldsymbol{V_l}$ and right robotic arm working space $\boldsymbol{V_r}$ are added and subtracted from the overlapping space $\boldsymbol{V_k}$, which is the actual working space of the two arms $\boldsymbol{V}$.

$$X = \frac{x}{b}, Y = \frac{y}{a}, Z = \frac{z}{c} \tag{3}$$

$$V_l = V_r = abc \iiint (X^2 + Y^2 + Z^2) dXdYdZ = \int_0^L dz \iint f(X, Y, Z) dXdY \tag{4}$$

$$l_k = \frac{4a - L_w}{2} \tag{5}$$

$$V_k = 2\int_{4a - L_w}^{a} dx \iint f(X, Y, Z) dYdZ \tag{6}$$

$$V = 2\int_0^L dz \iint f(X, Y, Z) dXdY - 2\int_{4a - L_w}^{a} dx \iint f(X, Y, Z) dYdZ \tag{7}$$

where $\boldsymbol{V}$ is the overlapping part of the dual-arm working space and grape-growing space, $\boldsymbol{W}$ is the grape-growing space length, $\boldsymbol{K}$ is the grape-growing space width, $\boldsymbol{L}$ is the grape-growing space height, $\boldsymbol{l_T}$ is the dual-arm mounting horizontal spacing, $\boldsymbol{H}$ is the height of the arm from the ground, $\boldsymbol{h_a}$ is the height of the arm from the grape-growing space, $\boldsymbol{S_{lt}}$ is the top area of the left arm working space and grape-growing space, $\boldsymbol{S_{lb}}$ is the bottom area of the left arm working

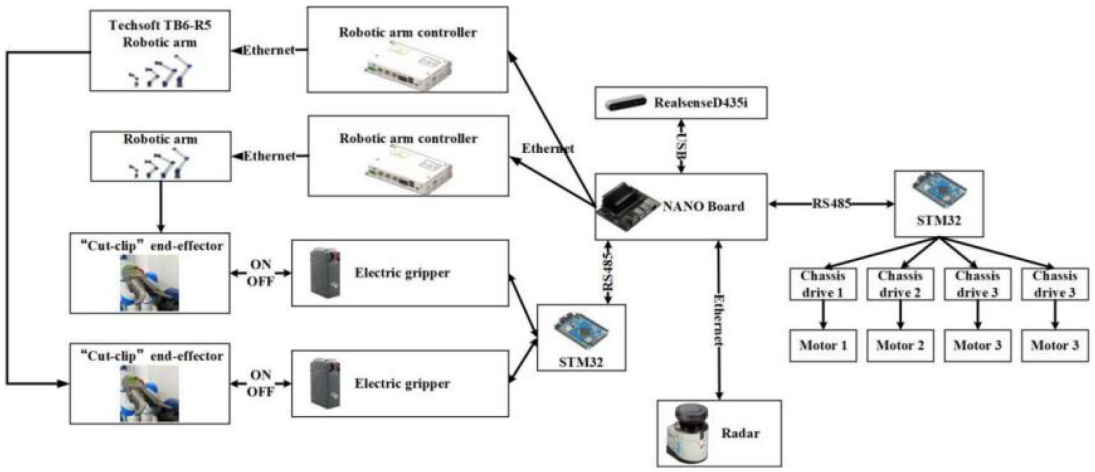


FIGURE 3
Hardware communication method.

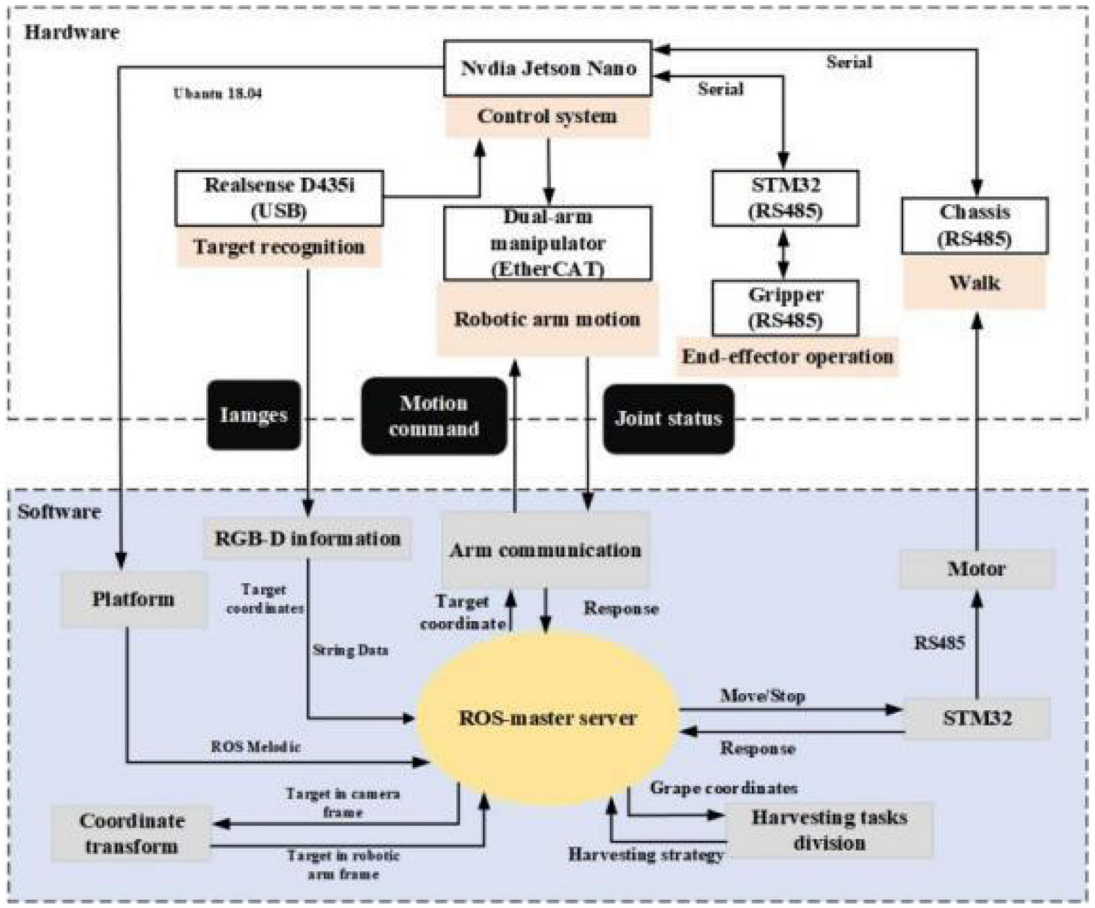


FIGURE 4
Control system of the dual-arm rapid harvesting robot.

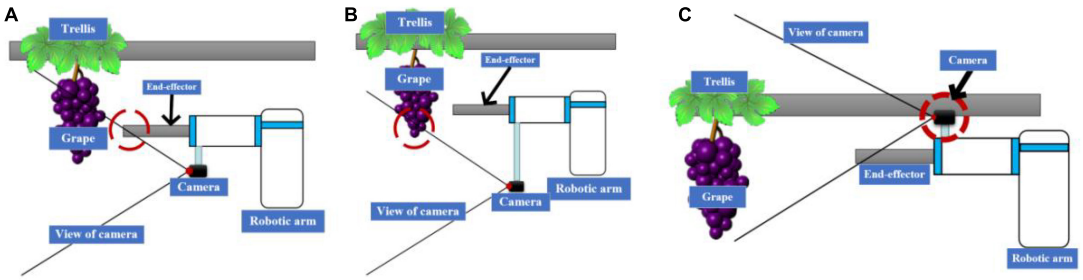


FIGURE 5
(A) Camera placed a short distance below. **(B)** Camera placed below at a distance. **(C)** Camera mounted above.

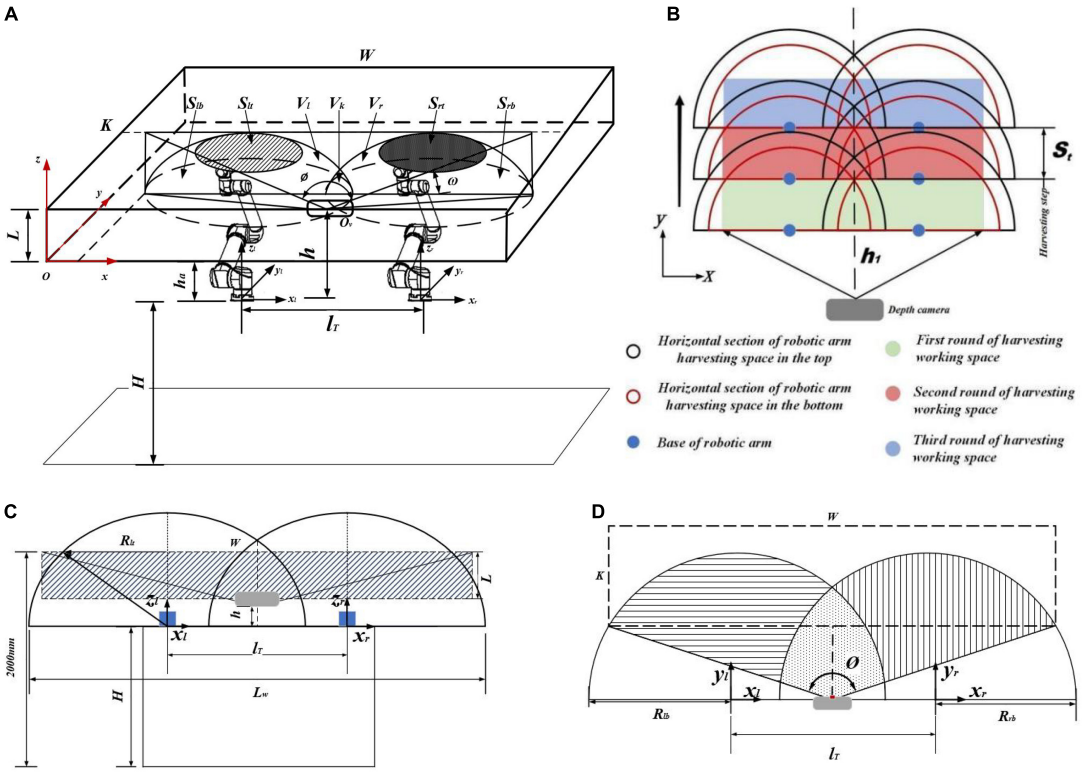


FIGURE 6
(A) Analysis of the "one eye-dual arm" working space and camera field of view. **(B)** Dual-arm working space and robot harvesting step. **(C)** Operating effect in the *XOZ* coordinate system. **(D)** Operating effect in the *XOY* coordinate system.

space and grape-growing space, S_{rt} is the top area of the right arm working space and grape-growing space, S_{rb} is the bottom area of the right arm working space and grape-growing space, and L_w is the working width of the two arms. O_v is the camera mounting position, h is the camera mounting height.

The area where the dual-arm workspace and camera field of view overlap is the area of harvesting that the robot can identify and harvest. Unreasonable arrangement of robot harvesting steps can effectively reduce the harvesting efficiency in grape-growing space and increase the number of missed grape targets. As shown in Figure 6B, the width of the camera field of view needs to be greater than the width of harvesting space W. The camera field of view takes the camera as the vertex. The directions of the FOV angle are extended. The shape of view is similar to a quadrilateral cone. By calculating the camera FOV angle, the camera field of view equation is derived. Threshold segmentation of the camera field of view effectively limits the range of the camera shot and filters interference.

From Figures 6B–D, we established the camera field-of-view equations. Relevant parameter constraints were established

based on the horizontal trellis, camera field-of-view range, and dual-arm working space.

Camera field of view space:

$$\pm \frac{x_o}{tan\frac{\emptyset}{2}} \pm \frac{z_o}{tan\frac{\omega}{2}} - 2y_o = 0 \quad (8)$$

where $\emptyset$ is the camera shooting horizontal field-of-view angle, ω is the camera shooting vertical field-of-view angle, (x_o, y_o, z_o) is the coordinate of the target point.

$$\begin{cases} 200mm \leq d_{view} \leq h_1 S_t \\ l_r + R_{lb} + R_{rb} \geq W \\ H + h + d_{height} \geq 2000mm \end{cases} \quad (9)$$

where d_{view} is the camera depth threshold range, d_{height} is the height of the camera field, S_t is the harvesting step of the robot, h_1 is the distance between the camera and robotic arms in the y direction, $\mathbf{l}_r$ is the mounting distance between dual arms, R_{lb} is the minimum working width of the left robotic arm in grape growing space, $\mathbf{R}_{rb}$ is the minimum working width of the right robotic arm in grape growing space.

By combining the characteristics of the horizontal trellis, camera field of view, and dual-arm working space, we obtain a reasonable installation position relationship between the camera and two arms:

$$\begin{cases} S_t = 800mm \\ H = 1400mm \\ h = 300mm \\ l_r = 1100mm \\ h_1 = 250mm \end{cases} \quad (10)$$

As shown in Figure 7, after MATLAB with Solidworks simulation, the "one eye-dual arm" structure ensures that the ends of the robot arm have sufficient space to move the trellis boundary so that the ends of the robot arm can reach the farthest end of the horizontal scaffolding in a flexible posture to complete harvesting operations, and the robot can be made in a harmonious proportion similar to the human form configuration without a lack of design aesthetics.

Large error tolerance of the "hand-eye" combination

(1) Rapid identification of multiple targets in one image.

The images of grapes inside a horizontal trellis obtained by RealSense D435i often exhibit multiple clusters of grapes. If each bunch of grapes needs to be identified once by the camera, it would greatly increase the harvesting time. To achieve rapid harvesting of multiple bunches of grapes in a horizontal trellis, it is necessary to achieve rapid identification of multiple bunches of grapes within an image.

Multiple bunches of grape bunches were often targeted in the images of grapes inside the horizontal trellis obtained by RealSense D435i. The camera directly acquires the depth values of all grapes in the image. According to Figure 8, we can calculate the camera depth threshold range as follows:

$$100mm < d < S_t + h_1 \quad (11)$$

When grape clusters are less than 100 mm from the camera, the camera cannot focus on the grapes. When grape clusters exceed the depth threshold, grape targets are beyond the working space of the two arms. As shown in Figure 8, four grape clusters were present in the image. The fourth grape cluster was cleared as background because the depth value exceeded the depth threshold. The other three grape clusters were harvested based on the depth value from smallest to largest.

(2) Fuzzy prediction of grape stem-cutting points based on grape contours.

Grape leaves, stems that are non-grapes, and grape clusters in trellises can interfere with the target stem identification in traditional algorithms. However, this fuzzy algorithm does not rely on the precise identification of grape stems. The algorithm constructs an external rectangle of grapes by HSV thresholds acquiring their geometric contours. The center of the external rectangle is the center of the grape profile in this algorithm and moves upward to speculate the coordinates of grape stems. When there is a small amount of cover in grape clusters, this algorithm can still quickly determine the inference of grape stem coordinates (Figure 9).

Grape contours were bounded to obtain parameter spike length and width values, and the center of grape coordinates $O(x_0, y_0, z_0)$ was calculated based on the distribution area. z_0 is the depth value of the center of the grape coordinate from the camera and can be obtained directly through the depth camera. The spatial coordinates of the grape-harvesting point $K(x_k, y_k, z_k)$ are calculated as follows:

$$\begin{cases} x_k = x_0 \\ y_k = y_0 + \frac{H}{2} + h \\ z_k = z_0 \end{cases} \quad (12)$$

(3) "Cut-clip" end effector for grape horizontal trellises.

Traditional finger end effectors often damage grapes at the finger end during the grape harvesting process. For Kyoho grapes, the stalks can reach 15 mm in diameter, and the weight of a single cluster can reach 400 g. RealSense D435i extracts the geometric contours of grape clusters to infer the calculation of grape stem-cutting points with visual recognition errors. Due to hand-eye calibration and coordinate conversion, the movement of the robotic arm has some motion errors. Both of these errors are generated by the design principle and algorithm. They are difficult to reduce or minimize. The end effector must grip the whole grape cluster when cutting the stem. It needs to be

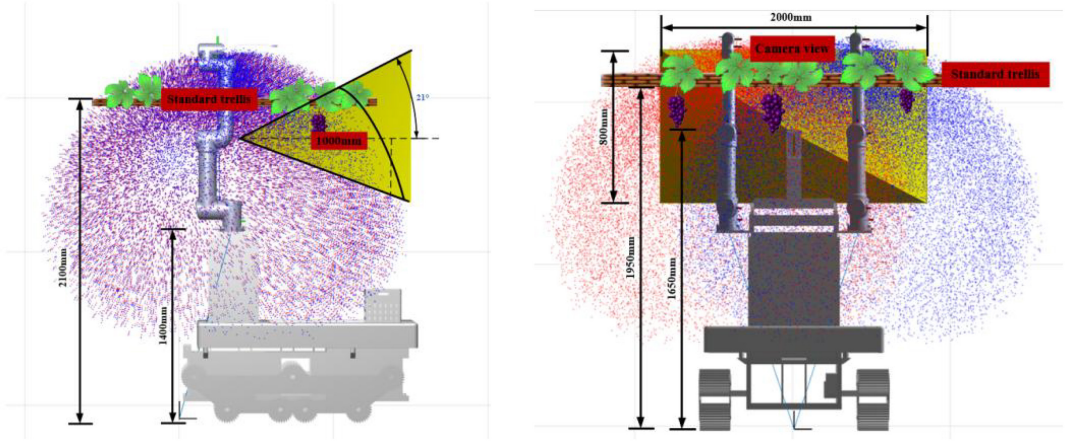


FIGURE 7
Operating space point cloud of the dual-arm harvesting robot.

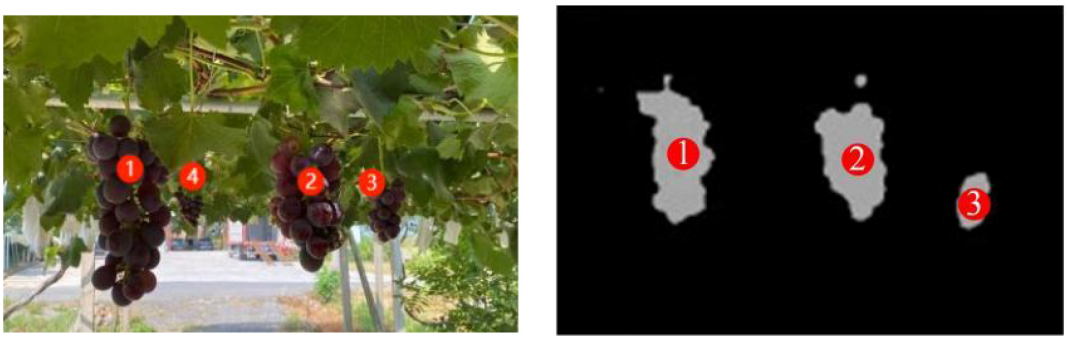


FIGURE 8
Identification of multiple bunches of grape clusters within one image.

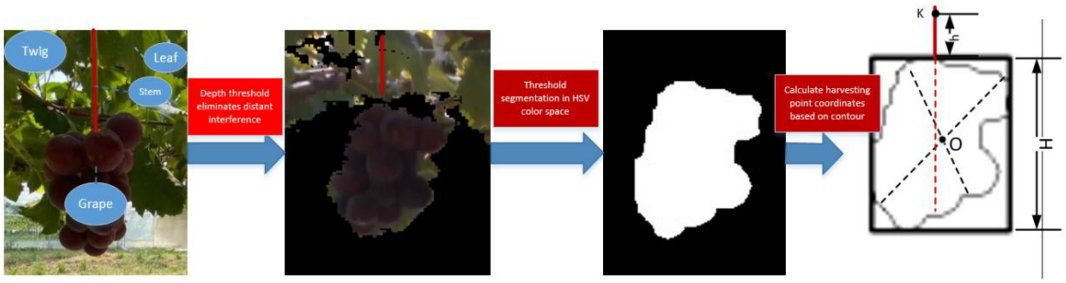


FIGURE 9
Calculation of the stem-cutting point.

transported from the standard trellis to the fruit box smoothly to ensure no damage.

Facing these requirements in grape harvesting, our research group has designed an end effector for rapid grape harvesting. The finger of the end effector is designed with certain curved angles. When the end effector begins to harvest, the finger with curved angles can reduce the negative effect of visual recognition errors and arm motion errors (Figures 10A,B). This structure enhances error tolerance in the x and y directions. It turns a harvesting point into a harvesting range.

Motors are used to control the fingers to open and close. Three sets of blades are mounted on the fingers. When the fingers are closed, the blades finish cutting the grape stems. At the same time, the lower part of the fingers is fitted

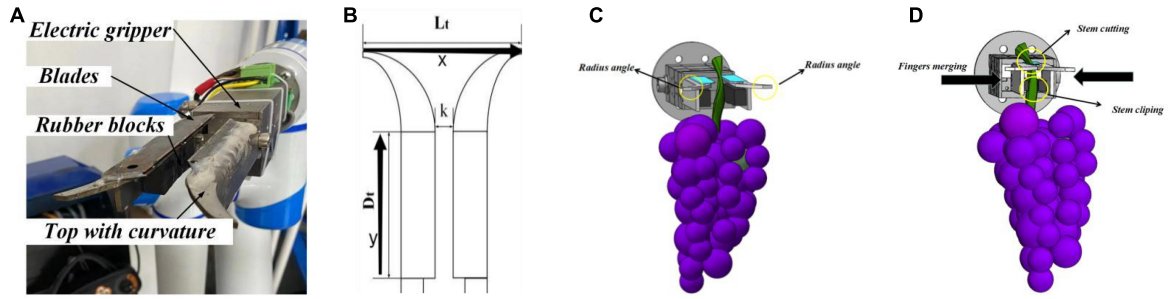


FIGURE 10
(A) Construction of the end effector; **(B)** error tolerance in the end effector; **(C)** end effector opening; and **(D)** end effector closing.

with a rubber block to hold grape clusters to cut the stems (Figure 10C). The end effector is simple in structure and only requires ± signals to complete the control process (Figure 10D). It enables the integration of cut-clip multitasking in rapid grape-harvesting tasks and significantly improves the harvesting efficiency and success rate.

(4) Error tolerant combination of the end effector and vision algorithms.

Nonstructural features exist within the horizontal trellis. Images contain not only grape clusters but also branches, leaves, the trellis, and another environment. It is a challenge to quickly acquire grape-harvesting points from complex backgrounds. When grape stems are obscured, interlaced, or overlapped, the stem-cutting point error is large. This leads to chaotic robotic arm movements, harvesting failures, and serious collision problems.

Faced with the special grape-harvesting requirement, our group obtained the coordinates of the center of the external rectangle based on the grape geometric profile and thus achieved the vertical upward prediction of grape stem-cutting points. By using the external rectangle of the grape cluster to predict stem-cutting points, even if grapes are partially obscured by the outline, the stem-cutting points can be predicted by the external rectangle with little error (Figure 11).

By the fuzzy prediction method for the grape stem-cutting points, grape clusters in horizontal trellises can be quickly recognized. When grape clusters were partially shaded, the external rectangle of grape clusters did not change significantly. Stem-cutting points were predicted by the external rectangle of grapes to reduce the errors of grape clusters in images. When the fingers of the end effectors have curved angles, the error of the stem-cutting points in the horizontal direction can be enhanced. As shown in Figure 10B, the design of the end effector produces some horizontal error tolerance degree $\boldsymbol{L}_t$ and depth error tolerance degree $\boldsymbol{D}_t$. It expands the original visually identified point into an area and improves the harvesting success rate.

It guides the stem to the area where the blade will cut. By the end effector mechanism, the point of the grape stem can be converted into an area range. This tolerable error method that combines software and hardware has significantly increased the success rate and harvesting efficiency of grapes in horizontal trellises.

Dual-arm harvesting strategies in the horizontal trellis

(1) Based on depth value "symmetric space segmentation" harvesting sequence.

Dual arms are not just a superposition of the operational efficiency of two robotic arms. The disorderly and random distribution of grapes on horizontal trellises means that the harvesting sequence and path for robotic arm harvesting operations need to be planned (Takano et al., 2019). Grape clusters captured by the camera view become harvesting targets, and the center axis plane of the camera field is used as the operation space segmentation reference plane. We divided the camera view into left working space and right working space based on the "symmetric space segmentation" method. Finally, the coordinating information of target grapes is sorted based on the depth values and transferred to the Cartesian coordinate system. When the target grape coordinate $x < 0$, the harvesting task is divided into the left arm workspace, and when $x > 0$, it is assigned to the right arm workspace. This ensures independent parallel operation between two robotic arms without interference and joint collision (Figure 12).

From Figure 12, multiple bunches of grapes L_1-L_3 and R_1-R_3 were found in the field of view of the camera in harvesting cycle 1. Dual arms harvest the target grape in their respective areas until all of them are harvested. When there are no grape targets in the camera field of view, the chassis will automatically run into harvesting cycle 2 and will harvest grape clusters L_4–L_6 and R_4-R_6.

(2) Danger and safety areas for dual-arm operation.

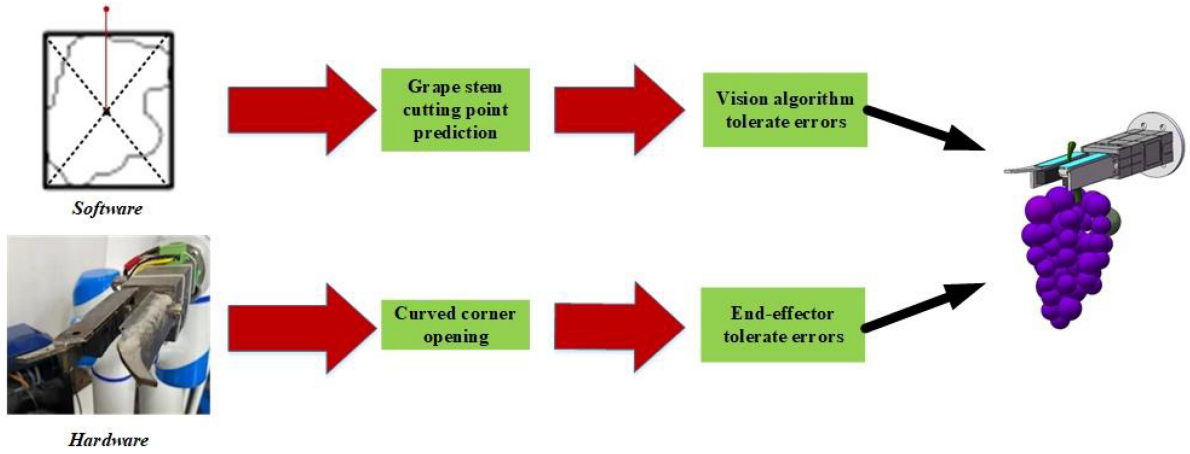


FIGURE 11
Error tolerance combinations of the end effector and vision algorithms.

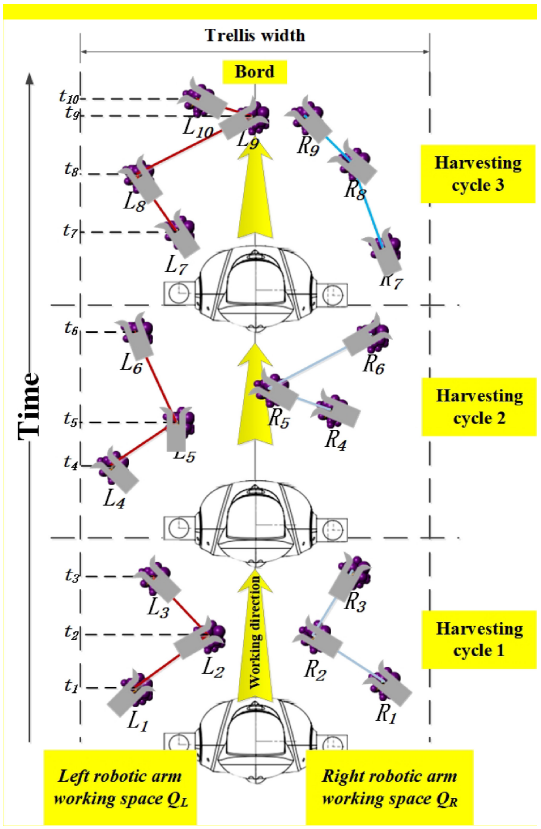


FIGURE 12
Dual-arm harvesting sequence in a horizontal trellis.

Robotic arms are used as electrical devices with independent control centers. Dual arms may be prone to collision and even serious damage. Therefore, we defined a dual-arm operating space and established a danger area and safe area in the working space. The fixed area in yellow shown in Figure 13 can be named the danger area. This means that we need to perform two scenario analyses:

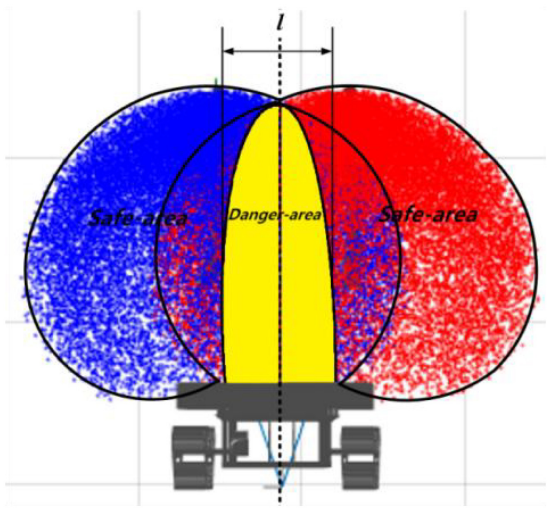


FIGURE 13
Danger area and safe area in dual-arm operating space.

(1) When grape targets are present in the safe area, the two arms do not collide. Therefore, the two arms can harvest independently and speedily without the need to restrict harvesting movement. Therefore, in this case, there is no need to change the strategy.
(2) If multiple bunches of grapes are present in the danger area, how to ensure that robotic arms can still complete the rapid harvesting process without collision must be determined.

TABLE 1 Different strategies for grape targets in different areas.

Scene	Use this strategy	Harvesting path
The two arms are out of the danger area	×	There is no risk of collision of the two arms, and independent and rapid harvesting can be achieved.
One arm in the danger area, the other arm in the safety area	√	The one arm in the danger area is treated as the master arm and has the priority of harvesting. The other arm needs to wait to complete its harvesting action before it starts moving.
The two arms are in the danger area	√	When grape targets are in the danger area, the harvesting priority of the arm needs to be determined based on the grape-harvesting order. The robotic arm with the priority will become the master arm. It will enter the danger area to harvest first.

(3) Asynchronous master–slave dual-robotic arm anticollision harvesting strategy in the danger area.

The danger area is a smaller part of the operating space of the dual arms. When multiple bunches of grape clusters are growing in the danger area, the movement of the two arms requires very careful planning. Otherwise, arm joints or end effectors are prone to collision. Our group proposes an asynchronous master–slave dual-robotic arm anticollision harvesting strategy in the danger area. This strategy is based on the conditional judgment of the grape cluster distribution location, as shown in Table 1.

The danger area occupies only a small part of the working space of the two arms. Therefore, the probability of this strategy being employed by two arms tends to be small, which does ensure the safety of robotic arms in harvesting work. As shown in Figure 14A, when multiple bunches of grapes are present in the danger area, the dual-arm strategy will be used for safe and rapid harvesting.

From Figure 14B, three bunches of grapes are in the danger area. Grape-harvesting tasks are divided into the left arm for bunches 1 and 3 and the right arm for bunch 2. At this time, the left arm acts as the master robotic arm, and it has the priority of harvesting the danger area. The right arm is a slave arm, and it needs to wait for the signal that the master arm has finished harvesting tasks. Then, it starts its harvesting mission.

From Figure 14C, there was a shift in the master–slave relationship between the two arms. The left arm moved down and out of the danger area. It was transformed from a master robot to a slave arm. Meanwhile, the right arm moved toward bunch 2 of the grape cluster in the danger area. It had priority access to the danger area for harvesting. The two arms entered the danger area for harvesting at different times. There is a time gap between the two arms in harvesting work.

Figures 14D–G show that converting the master–slave relationship between the two arms can ensure that the two arms work independently and smoothly in the danger area. When the robot adopts this strategy, it can reasonably use the time difference and robotic arm movement motion position in space. The high-speed harvesting work of the two arms in the danger area is an extremely difficult and complex task. In agricultural non-structural environments, a reasonable motion strategy for two arms often leads to great safety and efficiency improvements in the robot.

As shown in Figure 14H, if there is no harvesting target in the camera field of view, the robot will move some distance forward. A new harvesting cycle will start.

Experiments

Materials and methods

To verify the accuracy of large tolerance of the "hand-eye" combination and robot performance, both trellis and lab experiments were carried out:

(1) To test the accuracy and efficiency of the robotic arm in harvesting operations, we acquired the experimental errors in the harvesting process. An experimental platform was designed and built to finish grape cluster harvesting in a room. The grape-harvesting accuracy experimental platform is designed and produced, and two scale plates (0.5 m × 0.5 m) with a 2 mm grid size are combined and matched to form a coordinate experimental platform in Cartesian coordinates. The accuracy and performance of large error tolerance of the "hand-eye" combination were verified by every 30 harvesting experiments with shading grape clusters to different degrees (0–5, 6–20, and 21–40%). This platform can measure the coordinates of grape stem cutting point A by converting the robotic arm base coordinate system O_1 to the platform coordinate system O_2, and compare it with the visual recognition point and robotic arm motion point to derive a visual recognition accuracy error(mm) and arm positioning accuracy error(mm). Meanwhile, it needs to record harvesting time(s) and harvesting success rate (Figures 15A–D).
(2) Trellis performance experiments.

The experiments were conducted in September 2021 at the ErYa Vineyard in Jurong City, Jiangsu Province, China, where

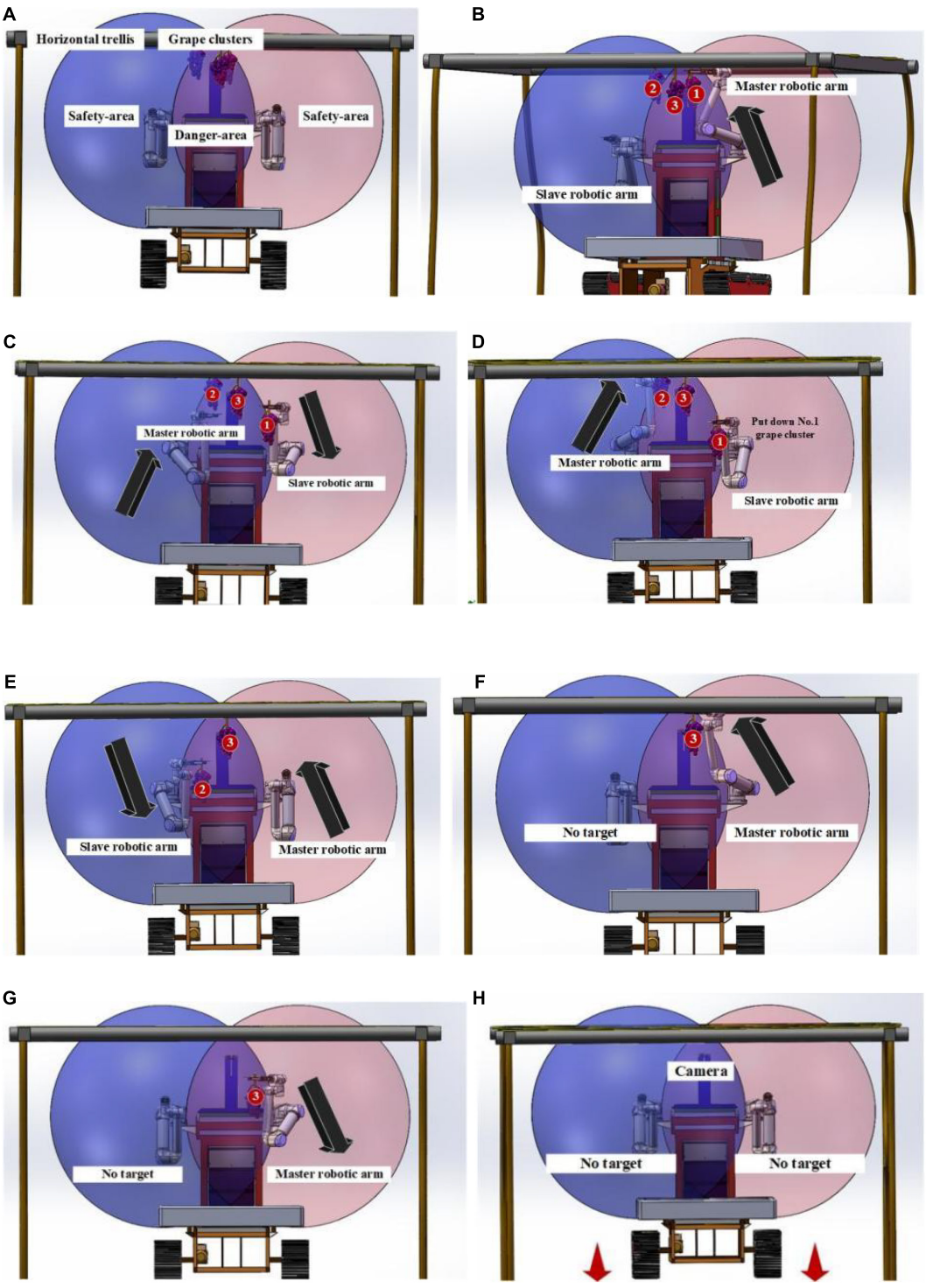


FIGURE 14
Asynchronous master–slave dual-robotic arm anticollision harvesting strategy in the danger area.

grapes were grown with a horizontal trellis type of cultivation. In this vineyard, grapes grew in good conditions, with most of the clusters hanging vertically below the trellis. The grape variety was Kyoho, which has large clusters, large grains, and purple–black fruits at maturity and is the main variety grown in grape production in China.With a horizontal trellis height

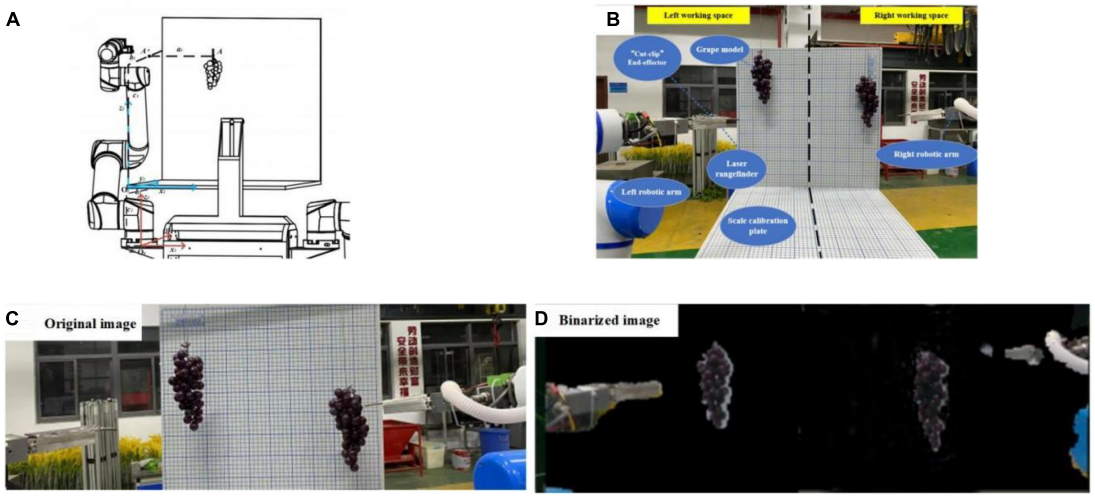


FIGURE 15
Figures of the indoor experiment: **(A)** harvesting point coordinate transformation, **(B)** indoor experimental schematic, **(C)** experimental original image, and **(D)** binarized image of grapes.

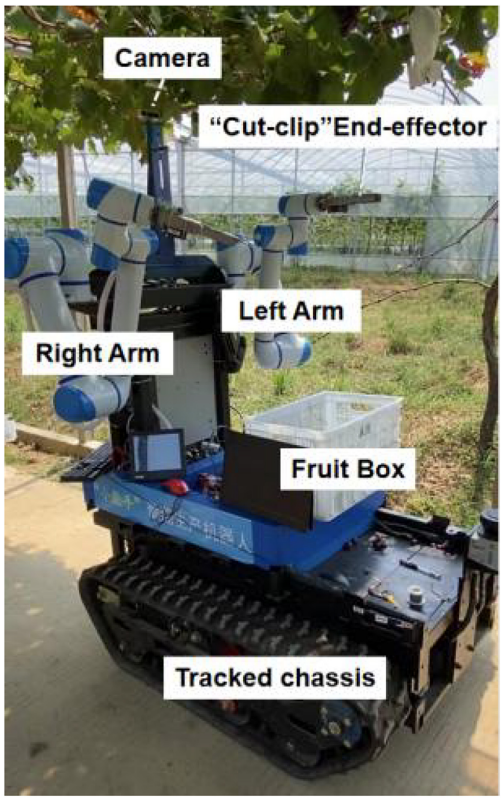


FIGURE 16
Dual-arm robot performing grape-harvesting operations.

of 2.0 m, a width of 2.5 m, and a trellis length of 30 m, this robot can meet the full range of coverage for harvesting in a single cycle inter row grape environment. There are no other obstacles around the grape-harvesting area, which can ensure that no exogenous emergency stopping occurs during the operation of the robot. The robotic harvesting process was captured in real time by the camera, recording the recognition success rate, harvesting success rate, and harvesting time of one grape cluster (Figure 16).

Results and discussion

(1) Lab experiments.

From Table 2, we know that the maximum visual recognition accuracy errors in the *x*, *y*, and *z* directions from the robotic arm base were 15.147, 13.689, and 16.330 mm, respectively, as the degree of obscuration of the grape bunches by the leaves increased, thus showing that the integrity of grape bunches' contours accounted for a great deal of the impact on the visual recognition accuracy of the camera. The motion positioning errors of the left and right robotic arms were 2.885, 3.972, and 2.715 mm and 2.471, 3.289, and 3.775 mm, respectively, indicating that these robotic arms were well positioned and could support end effectors in reaching the grape-harvesting point accurately. The above errors were adjusted by the structure and design of the end effector, which can be applied to the grape-harvesting accuracy requirements under the operating conditions of a horizontal trellis. The average single-cycle completion time is 8.45 s. To explore and optimize the visual recognition capability of this robot, harvest failure tests were analyzed.

From harvesting failure tests, the binarized images show grape cluster contours, and the grapes not obscured by the leaves are easily obtained as complete contours, allowing

TABLE 2 Accuracy experiment results.

Degree of leaf shade (%)	Visual recognition accuracy error (mm)	Arm positioning accuracy error (left arm) (mm)	Arm positioning accuracy error (right arm) (mm)	Grape harvesting time (s)	Success rate (%)
0–5	(10.899, 8.552, 6.337)	(2.098, 3.003, 3.539)	(2.964, 2.363, 2.086)	8.47	93.3
6–20	(11.502, 10.141, 12.639)	(3.497, 3.551, 2.314)	(3.443, 3.605, 5.203)	8.23	86.7
21–40	(15.147, 13.689, 16.330)	(3.060, 5.363, 2.292)	(3.414, 3.901, 4.036)	8.66	73.3

TABLE 3 Trellis performance experiment results.

Grape cluster ID/ number	Successful visual recognition	Successful harvest	One grape cluster harvesting time/s	Damaged grains/ number	Grape damage rate
1	√	√	8.14	0	0
2	√	√	8.76	0	0
3	√	√	8.60	0	0
4	√	√	8.93	0	0
5	√	√	9.43	0	0
6	×	×		–	
7	√	√	8.72	0	0
8	√	√	8.10	2	4.39%
9	×	×		–	
10	√	√	8.98	0	
11	√	√	9.25	1	2.02%
12	√	√	8.78	0	0
13	√	√	9.34	0	0
14	√	√	8.64	0	0
15	√	√	8.19	0	0
Average	86.7%	86.7%	8.76	0.23	

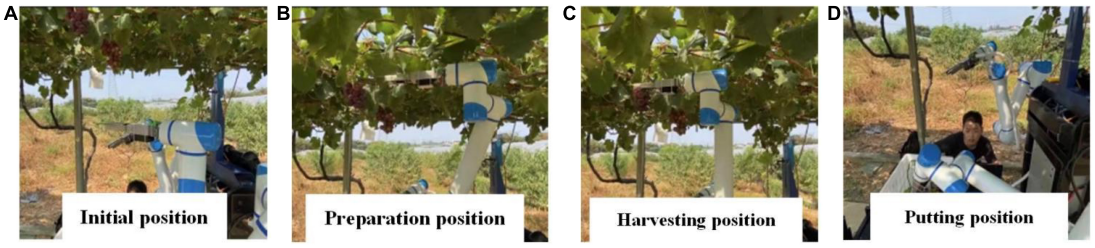


FIGURE 17
Robotic arm harvesting postures: **(A)** initial position; **(B)** harvesting preparation; position **(C)** harvesting position; and **(D)** putting position.

accurate calculation of the stem-cutting location. However, the grape area shrinks with increasing leaf occlusion resulting in many deviations in the center of the grape contour and the stem-cutting point coordinates. This affected the success rate of subsequent harvesting by robotic arms. After subsequent iterations and changes in test conditions, the factors affecting this phenomenon were identified.

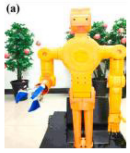
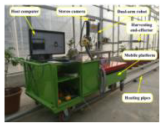
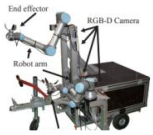
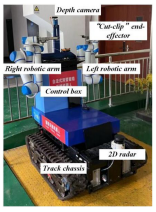
- Uneven light distribution.

Influenced by the sunlight irradiation direction and grape growth contour, the images captured by the camera were incomplete, with abnormalities such as mutilation, deformation, and overlap of grape clusters within the images, resulting in deviations in the generated grape stem-harvesting points. However, in a normal horizontal trellis, grape leaves and branches grow at the top of the trellis, and the sunlight intensity generally does not interfere greatly with camera recognition.

- Shaking of the grape model.

The selected grape model was made of plastic, with low weight and weak resistance to external interference, resulting in slight shaking during photography. Before the

TABLE 4 Comparison between dual arms robots.

References	Products	Harvesting type	Harvesting success rate (%)	Harvesting efficiency (amount/hour)	Scenes
Arad et al., 2020		Sweet pepper	61	24	Greenhouse
Yu et al., 2021		Apple	72	14.6	Indoor
Ling et al., 2019		Tomato	87.5	30	Greenhouse
SepúLveda et al., 2020		Aubergine	91.67	26	Indoor
Yoshida et al., 2022		Apple	/	10	Field
This research		Grape	86.7	8.76	Field

trellis performance test, lighting was installed on the head of this robot to reduce the interference of natural light on camera recognition. The grapes planted in the trellis were hung vertically from the top, and the average weight of each grape cluster was close to 400 g. Thus, they were highly resistant to external interference and therefore only slightly swayed, with minimal effect on camera recognition.

(2) Trellis experiments.

This grape-harvesting robot advanced to the horizontal trellis and started harvesting above this trellis with the "sequential mirroring" strategy based on the depth information. The grape damage rate is the mass of grape clusters from falls, breaks, and bruises as a percentage of the mass of all harvested grape clusters. The number of harvesting successes, single-cycle dual-arm harvesting time, and grape damage rate was used as the main indicators to measure the quality of the dual-arm grape-harvesting robot in the tests (Table 3).

- Results of visual identification.

The visual images of the dual-arm grape-harvesting robot show the fruit shape contour, segmented depth threshold, and grape target binarization image. The vision system calculates the image center of the grape-based on the binarization recognition image and derives the Cartesian spatial coordinate information of the grape stem-harvesting points. Among the 15 sets of experiments, 13 sets of experiments were completed. The fusion of depth information and color information determines the harvesting order arrangement, and the visual localization accuracy reaches 86.7% without neural network training, which can realize fast localization recognition in normal agricultural harvesting work.

- Continuous grape-harvesting test.

By analyzing the dynamics of each harvesting process and stem separation points of robotic arms, we analyzed the

displacement change relationship between the stem and grapes during the grape-harvesting process and verified the single journey continuous harvesting method from the initial position, harvesting preparation point, harvesting point, and grape-putting point. The data in Table 3 show that the success rate of harvesting is 86.7%, and the main reason for the failure is the small size of the grape, which affects the correct conversion of the final harvesting coordinates. After 13 successful harvesting tests, the average harvesting time of one grape cluster is 8.76 s. The operating speed of the robotic arm was only set to 40% of the maximum speed of joint motion, and the single-cycle operating efficiency was still greatly improved after the subsequent structural stabilization of the robot (Figures 17A–D).

And we compared it with some currently used fruit and vegetables harvesting dual arms robots as shown in Table 4.

By comparing these advanced harvesting dual arms robots, our robot has a faster harvesting efficiency, and reliable harvesting success rate and can be adapted to the complex vineyard.

Conclusion

As a multipurpose fruit that easily falls off or break, how to achieve rapid and undamaged harvesting of grapes has become an urgent problem for the current grape industry worldwide. In this study, a dual-arm grape-harvesting robot is developed based on grape-harvesting demand in the special growing environment of horizontal trellises. This robot accomplishes the fusion and extraction of spatial multitarget information by a single depth camera and simplifies the calculation of 3D graphic information into spatial point coordinates. A "one eye-to-dual hands" vision servo system is built, and a single RGB-D camera is used to divide the field of view into equal tasks for two robotic arms, to locate multiple grape targets quickly and continuously, and transmit the spatial information of grape harvesting to the corresponding two robotic arms based on the corresponding spatial growth position distribution. The whole process of rapidly harvesting grapes was completed by transforming visual information and digital information into robotic machine signals.

To simulate the real environment of grape harvesting in the horizontal trellis, 30 sets of positioning accuracy tests were conducted with different degrees of leaf shading. Without neural network training, when the degree of leaf shading was 0–5%, the harvesting success rate was 93.3%, and one grape cluster harvesting time was 8.47 s. When the degree of leaf shading was 6–20%, the harvesting success rate was 86.7%, and one grape cluster harvesting time was 8.23 s. When the degree of leaf shading was 21–40%, the harvesting success rate was 73.3%, and one grape cluster harvesting time was 8.66 s, which met the requirements of rapid location identification and low-loss harvesting of grape clusters in a real horizontal trellis environment. After the trellis performance harvesting test, out of the 15 sets of experiments, 13 sets of experiments were completed with accurate identification, the visual positioning accuracy reached 86.7%, and the average harvesting time of one grape cluster was 8.76 s without neural network training, so fast positioning identification and rapid low-loss harvesting of grape clusters were achieved in a real horizontal scaffolding environment.

In the next step, because grape harvesting is still not faster than human harvesting, we will continue to work on optimizing all aspects of the robotic arm harvesting motion process. At the same time, we will conduct research on the minimization of robotic arm motion paths in nonstructural environments. The work of two arms in grape harvesting still holds great promise for research. All the technical details will be reported in the next study.

Data availability statement

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

Author contributions

YJ and JL: conceptualization and validation. YJ, JL, and JW: methodology. YJ: software. JW, WL, HS, and YP: formal analysis. YJ, WL, and YP: investigation. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Natural Science Foundation of China (NSFC) (No. 31971795), Priority Academic Program Development of Jiangsu Higher Education Institutions (No. PAPD-2018-87), and Project of Faculty of Agricultural Equipment of Jiangsu University (No. 4111680002).

Acknowledgments

Appreciations are given to the editor and reviewers of the Journal.

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

References

Arad, B., Balendonck, J., Barth, R., Ben-Shahar, O., Edan, Y., Hellström, T., et al. (2020). Development of a sweet pepper harvesting robot. *J. Field Robot.* 37, 1027–1039.

Barth, R., Hemming, J., and van Henten, E. J. (2016). Design of an eye-in-hand sensing and servo control framework for harvesting robotics in dense vegetation. *Biosyst. Eng.* 146, 71–84. doi: 10.1016/j.biosystemseng.2015.12.001

Chen, M., Tang, Y., Zou, X., Huang, Z., Zhou, H., and Chen, S. (2021). 3D global mapping of large-scale unstructured orchard integrating eye-in-hand stereo vision and SLAM. *Comput. Electr. Agricult.* 187:106237. doi: 10.1016/j.compag.2021.106237

Isrigova, T. A., Salmanov, M. M., Makuev, G. A., Saipullaeva, A., Isrigova, V. S., Ramazanov, S. R., et al. (2021). "Studying the crop yield influence on the commercial quality of the promising table grapes," in *Proceedings of the IOP Conference Series: Earth and Environmental Science*, Vol. 624, (Bristol: IOP Publishing), 012162. doi: 10.1088/1755-1315/624/1/012162

Jianying, F., Xia, W., Zetian, F., and Weisong, M. (2014). Assessment of consumers' perception and cognition toward table grape consumption in China. *Br. Food J.* 116, 611–628. doi: 10.1108/BFJ-04-2012-0101

Kalampokas, T, Vrochidou, E, Papakostas, G. A., Pachidis, T., and Kaburlasos, V. G. (2021). Grape stem detection using regression convolutional neural networks. *Comput. Electr. Agricult.* 186:106220. doi: 10.1016/j.compag.2021.106220

Kootstra, G., Wang, X., Blok, P. M., Hemming, J., and Van Henten, E. (2021). Selective harvesting robotics: current research, trends, and future directions. *Curr. Robot. Rep.* 2, 95–104. doi: 10.1007/s43154-020-00034-1

Kurtser, P., and Edan, Y. (2020). Planning the sequence of tasks for harvesting robots. *Robot. Autonomous Syst.* 131:103591. doi: 10.1016/j.robot.2020.103591

Levin, M., and Degani, A. (2019). A conceptual framework and optimization for a task-based modular harvesting manipulator. *Comput. Electr. Agricult.* 166:104987. doi: 10.1016/j.compag.2019.104987

Ling, X., Zhao, Y., Gong, L., Liu, C., and Wang, T. (2019). Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision. *Robot. Auton. Syst.* 114, 134–143. doi: 10.1016/j.robot.2019.01.019

Liu, J., Yuan, Y., Gao, Y., Tang, S., and Li, Z. (2019). Virtual model of grip-and-cut picking for simulation of vibration and falling of grape clusters. *Trans. ASABE* 62, 603–614. doi: 10.13031/trans.12875

Luo, L., Tang, Y., Zou, X., Ye, M., Feng, W., and Li, G. (2016). Vision-based extraction of spatial information in grape clusters for harvesting robots. *Biosyst. Eng.* 151, 90–104. doi: 10.1016/j.biosystemseng.2016.08.026

Majeed, Y., Karkee, M., Zhang, Q., Fu, L., and Whiting, M. D. (2021). Development and performance evaluation of a machine vision system and an integrated prototype for automated green shoot thinning in vineyards. *J. Field Robot.* 38, 898–916. doi: 10.1002/rob.22013

Mehta, S. S., MacKunis, W., and Burks, T. F. (2014). Nonlinear robust visual servo control for robotic citrus harvesting. *IFAC Proc. Vol.* 47, 8110–8115. doi: 10.3182/20140824-6-ZA-1003.02729

Organisation internationale de la vigne et du vin [OIV] (2020). *OIV Statistical Report on World Vitiviniculture*. Paris: OIV. doi: <doi>

Ozkan, B., Fert, C., and Karadeniz, C. F. (2007). Energy and cost analysis for greenhouse and open-field grape production. *Energy* 32, 1500–1504.

doi: 10.1016/j.energy.2006.09.010

Peng, Y., Zhao, S., and Liu, J. (2021). Segmentation of overlapping grape clusters based on the depth region growing method. *Electronics* 10:2813. doi: 10.3390/electronics10222813

Piazzolla, F., Pati, S., Amodio, M. L., and Colelli, G. (2016). Effect of harvest time on table grape quality during on-vine storage. *J. Sci. Food Agricult.* 96, 131–139. doi: 10.1002/jsfa.7072

Possingham, J. V. (2006). "Developments in the production of table grapes, wine and raisins in tropical regions of the world," in *Proceedings of the International Symposium on Grape Production and Processing*, Vol. 785, (Maharashtra), 45–50. doi: 10.17660/ActaHortic.2008.785.1

Seol, J., Lee, S., and Son, H. I. (2020). A review of end-effector for fruit and vegetable harvesting robot. *J. Korea Robot. Soc.* 15, 91–99. doi: 10.7746/jkros.2020.15.2.091

SepúLveda, D., Fernández, R., Navas, E., Armada, M., and González-De-Santos, P. (2020). Robotic aubergine harvesting using dual-arm manipulation. *IEEE Access* 8, 121889–121904.

Seyyedhasani, H., Peng, C., Jang, W. J., and Vougioukas, S. G. (2020). Collaboration of human pickers and crop-transporting robots during harvesting–Part I: Model and simulator development. *Comput. Electr. Agricult.* 172:105324. doi: 10.1016/j.compag.2020.105324

Suvoéarev, K., Blanco, O., Faci, J. M., Medina, E. T., and Martínez-Cob, A. (2013). Transpiration of table grape (Vitis vinifera L.) trained on an overhead trellis system under netting. *Irrigat. Sci.* 31, 1289–1302. doi: 10.1007/s00271-013-0404-2

Takano, W., Takahashi, T., and Nakamura, Y. (2019). Sequential Monte Carlo controller that integrates physical consistency and motion knowledge. *Auton. Robots* 43, 1523–1536. doi: 10.1007/s10514-018-9815-5

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.* 11:510. doi: 10.3389/fpls.2020.00510

Vrochidou, E., Bazinas, C., Manios, M., Papakostas, G. A., Pachidis, T. P., and Kaburlasos, V. G. (2021). Machine vision for ripeness estimation in viticulture automation. *Horticulturae* 7:282. doi: 10.3390/horticulturae7090282

Wang, Y., Yang, Y., Yang, C., Zhao, H., Chen, G., Zhang, Z., et al. (2019). End-effector with a bite mode for harvesting citrus fruit in random stalk orientation environment. *Comput. Electr. Agricult.* 157, 454–470. doi: 10.1016/j.compag.2019.01.015

Wang, Z., Zhou, J., Xu, X., Perl, A., Chen, S., and Ma, H. (2017). Adoption of table grape cultivars: An attribute preference study on Chinese grape growers. *Sci. Horticult.* 216, 66–75. doi: 10.1016/j.scienta.2017.01.001

Williams, L. E., and Fidelibus, M. W. (2016). Measured and estimated water use and crop coefficients of grapevines trained to overhead trellis systems in California's San Joaquin Valley. *Irrigat. Sci.* 34, 431–441. doi: 10.1007/s00271-016-0513-9

Yoshida, T., Onishi, Y., Kawahara, T., and Fukao, T. (2022). Automated harvesting by a dual-arm fruit harvesting robot. *Res. Square [Preprint]*. doi: 10.21203/rs.3.rs-1455100/v1

Yu, X., Fan, Z., Wang, X., Wan, H., Wang, P., Zeng, X., et al. (2021). A lab-customized autonomous humanoid apple harvesting robot. *Comput. Electr. Eng.* 96:107459. doi: 10.1016/j.compeleceng.2021.107459

Zhao, Y., Gong, L., Liu, C., and Huang, Y. (2016b). Dual-arm robot design and testing for harvesting tomato in greenhouse. *IFAC-PapersOnLine* 49, 161–165. doi: 10.1016/j.ifacol.2016.10.030

Zhao, Y., Gong, L., Huang, Y., and Liu, C. (2016a). A review of key techniques of vision-based control for harvesting robot. *Comput. Electr. Agricult.* 127, 311–323. doi: 10.1016/j.compag.2016.06.022

Zhu, D. S., Wu, X. F., Zhang, Q. M., Meng, X. J., and Li, J. R. (2014). "Rapid analysis of table grape maturity based on color difference," in *Applied Mechanics and Materials*, Vol. 448, eds V. Adimule, et al. (Switzerland: Trans Tech Publications Ltd), 1133–1137. doi: 10.4028/www.scientific.net/AMM.448-453.1133

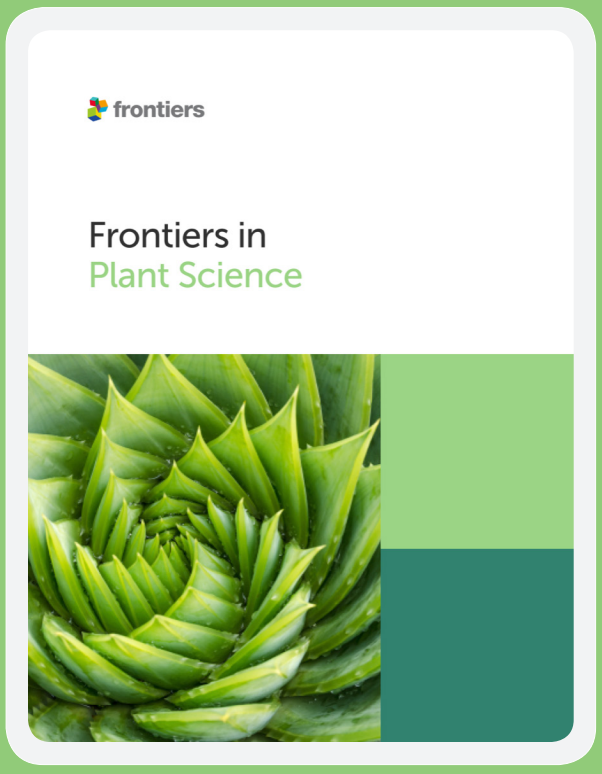
Frontiers in
Plant Science

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