Investigating AI-based smart precision agriculture techniques

Edited by
Uzair Aslam Bhatti and Mehedi Masud

Published in
Frontiers in Plant Science

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

Investigating AI-based smart precision agriculture techniques

Topic editors

Uzair Aslam Bhatti — Hainan University, China

Mehedi Masud — Taif University, Saudi Arabia

Topic Coordinator

Sibghat Ullah Bazai — Balochistan University of Information Technology, Engineering and Management Sciences, Pakistan

Table of
contents

Editorial: Investigating AI-based smart precision agriculture techniques

Uzair Aslam Bhatti[1]*, Mehdi Masud[2], Sibghat Ullah Bazai[3] and Hao Tang[1]*

[1]School of Information and Communication Engineering, Hainan University, Haikou, China, [2]Department of Computer Science, Taif University, Taif, Saudi Arabia, [3]Balochistan University of Information Technology, Engineering, and Management Sciences (BUITEMS), Quetta, Pakistan

Editorial on the Research Topic
Investigating AI-based smart precision agriculture techniques

Overview

The monsoon plays a pivotal role in determining agricultural output. The success of crops planted during a monsoon season is highly contingent on the prevailing weather conditions. Access to real-time meteorological information is crucial for farmers to make informed decisions regarding crop management, thereby reducing the risks and losses associated with adverse climatic conditions.

Agricultural fields are typically managed on a field-by-field basis, without considering the spatial and temporal variability of the soil. This approach can lead to uninformed decisions by farmers regarding inputs such as fertilizers, irrigation systems, and labor, resulting in suboptimal harvests. To address this, precision agriculture aims to optimize resource utilization by effectively managing the temporal and geographical variability of soil and ecosystem conditions. The advent of the Internet of Things (IoT) and sensor-edge connecting devices has greatly facilitated the collection of agricultural data in a smart manner for farmers.

In addition to weather-related challenges, economic difficulties also impact farming practices and productivity, particularly in rural and semi-rural areas. Farmers face daily obstacles such as pesticide use, water scarcity, resource limitations, and poor soil quality. Overcoming these challenges requires the strategic determination of best practices and approaches.

Smart precision agriculture emerges as an innovative solution that leverages cutting-edge technology to enhance crop yields sustainably. By integrating smart IoT devices and sensors, farmers can optimize agricultural output while minimizing their field work time. Smart technologies enable more efficient resource usage, including reduced water and power consumption, and constant monitoring of variables like humidity and temperature. Internet of Things-based smart farming utilizes multiple sensors, measuring parameters such as humidity, temperature, and soil moisture, to monitor field conditions effectively.

Despite the potential benefits, Smart Sustainable Agriculture (SSA) faces challenges due to insufficient investment in research and development. Additionally, complex barriers arise from the fragmented nature of agricultural processes, encompassing aspects such as the management and operation of IoT/AI machines, remote sensing, environmental impact assessment, data sharing and management, interoperability, and the analysis and storage of extensive datasets.

The provided text discusses various research papers and studies related to AI-based smart precision agriculture techniques. Here is a summary of each study:

Chinese agricultural named entity recognition

This study focuses on improving named entity recognition in Chinese agricultural texts, specifically in the context of kiwifruit diseases and pests. The researchers propose a novel model called KIWINER, which incorporates new word detection, an attention-based softlexicon module, and a parallel connection criss-cross attention module. The model achieves high F1-scores on multiple datasets, demonstrating its effectiveness in recognizing kiwifruit-related named entities (Zhang et al.).

Infrared and visible image fusion in agriculture

The paper presents a distributed fusion architecture called RADFNet for combining infrared and visible images in agricultural applications. The architecture utilizes residual CNN, edge attention, and multiscale channel attention to improve image quality and eliminate environmental interference. Experimental results show that RADFNet outperforms existing image fusion algorithms in terms of visual effect and quantitative metrics (Feng et al.).

Dome-type planted pumpkin autonomous harvesting framework

This study introduces a framework for autonomous harvesting of dome-type planted pumpkins. The framework includes a keypoint detection method using instance segmentation architecture, combining transformer network and point rendering to address overlapping and improve segmenting precision. Experimental results on a pumpkin image dataset demonstrate the effectiveness of the proposed method in instance segmentation and keypoint detection, with promising application prospects in fruit picking tasks (Yan et al.).

Genetic diversity analysis of Hopea hainanensis

The research focuses on the genetic diversity of Hopea hainanensis, an endangered tree species found in Hainan Island,

China. Using SNP and genotyping-by-sequencing technology, the study analyzes the genetic diversity among different populations of Hopea hainanensis in fragmented habitats. The results reveal low genetic diversity, highlighting the need for genetic diversity research in the conservation of rare and endangered plants (Chen et al.).

Cassava leaf disease classification

This paper addresses the classification of cassava leaf diseases using deep convolutional neural networks. A multi-scale fusion model based on attention mechanism is proposed to enhance disease feature extraction from cassava leaves. The model achieves improved classification performance compared to the original model, providing support for the recognition and early diagnosis of plant disease leaves (Liu et al.).

Weed detection in turfgrass

The study focuses on weed detection in turfgrass using deep learning methods. Various convolutional neural networks (DenseNet, EfficientNet-v2, and ResNet) are trained to detect weeds susceptible to herbicides, enabling site-specific weed detection. The results demonstrate high F1 scores and MCC values for most weed species, except for those with similar plant morphology. The proposed method provides an effective strategy for precision herbicide application (Jin et al.).

Crop rotation and soil health

This research examines the impact of different vegetable cropping systems on soil chemical properties, eggplant photosynthesis, and antioxidant functioning. Leafy vegetable rotation systems are found to significantly improve soil organic matter and available nutrients, as well as enhance eggplant growth and yield. The rotation systems also lead to higher antioxidant enzyme activity, reducing oxidative damage to membranes. The study highlights the benefits of crop rotation for improving the growth and yield of eggplant (Ghani et al.).

Objective evaluation of turfgrass cultivars

The project addresses the subjectivity in the evaluation of turfgrass cultivars using ordinal data. A model-based approach is proposed to minimize subjectivity and enable objective comparisons of cultivars across different test locations. The model is fitted in a Bayesian framework, allowing the estimation of additional parameters and providing better separation of cultivar means. The approach improves the evaluation procedure and enables more realistic comparisons (Qu et al.).

Yolo optimization

YOLOv7 maize pests identification method incorporating the Adan optimizer is proposed for the timely and accurate detection of major pests of corn. The study focuses on three major corn pests: corn borer, armyworm, and bollworm. A corn pests dataset is constructed using data augmentation techniques to address the issue of limited pest data. The YOLOv7 network is chosen as the detection model, and the Adan optimizer is introduced to replace the original optimizer for improved efficiency and accuracy while reducing computational costs (Zhang et al.).

UNET+CBAM disease classification

This research focuses on the identification of apple diseases, specifically Alternaria blotch and brown spot diseases, aiming to improve production efficiency and quality. The paper proposes a disease spot segmentation and disease identification method based on DFL-UNet+CBAM. The primary issues addressed are the low recognition accuracy and poor performance of small spot segmentation in apple leaf disease recognition. The objective is to accurately prevent and control apple diseases, minimize fruit quality degradation, yield reduction, and associated economic losses. The proposed DFL-UNet+CBAM model incorporates a hybrid loss function comprising Dice Loss and Focal Loss (Zhang et al.).

Pepper leaf segmentation

The study focuses on segmenting pepper leaves from images to aid in the control of pepper leaf diseases. A bidirectional attention fusion network called BAF-Net is proposed (Zhang et al.).

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

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.

Check for updates

*CORRESPONDENCE
Dangdang Niu
niudd@nwafu.edu.cn
Hongming Zhang
zhm@nwafu.edu.cn

Lexicon and attention-based named entity recognition for kiwifruit diseases and pests: A Deep learning approach

Lilin Zhang[1], Xiaolin Nie[1], Mingmei Zhang[1], Mingyang Gu[1], Violette Geissen[2], Coen J. Ritsema[2], Dangdang Niu[1]* and Hongming Zhang[1]*

[1]College of Information Engineering, Northwest Agricultural and Forestry (A&F) University, Yangling, China, [2]Soil Physics and Land Management Group, Wageningen University, Wageningen, Netherlands

Named Entity Recognition (NER) is a crucial step in mining information from massive agricultural texts, which is required in the construction of many knowledge-based agricultural support systems, such as agricultural technology question answering systems. The vital domain characteristics of Chinese agricultural text cause the Chinese NER (CNER) in kiwifruit diseases and pests to suffer from the insensitivity of common word segmentation tools to kiwifruit-related texts and the feature extraction capability of the sequence encoding layer being challenged. In order to alleviate the above problems, effectively mine information from kiwifruit-related texts to provide support for agricultural support systems such as agricultural question answering systems, this study constructed a novel Chinese agricultural NER (CANER) model KIWINER by statistics-based new word detection and two novel modules, AttSoftlexicon (Criss-cross attention-based Softlexicon) and PCAT (Parallel connection criss-cross attention), proposed in this paper. Specifically, new words were detected to improve the adaptability of word segmentation tools to kiwifruit-related texts, thereby constructing a kiwifruit lexicon. The AttSoftlexicon integrates word information into the model and makes full use of the word information with the help of Criss-cross attention network (CCNet). And the PCAT improves the feature extraction ability of sequence encoding layer through CCNet and parallel connection structure. The performance of KIWINER was evaluated on four datasets, namely KIWID (Self-annotated), Boson, ClueNER, and People's Daily, which achieved optimal F_1-scores of 88.94%, 85.13%, 80.52%, and 92.82%, respectively. Experimental results in many aspects illustrated that methods proposed in this paper can effectively improve the recognition effect of kiwifruit diseases and pests named entities, especially for diseases and pests with strong domain characteristics

1 Introduction

Kiwifruit is one of the economic sources of the planting industry in China, but owing to the impact of pests and diseases, the overall level of kiwifruit quality in China is not high at present (Jiang and Zong, 2020). Chinese named entity recognition in the field of agriculture aims to recognize the boundaries and categories of agriculture-related entities from unstructured agricultural texts, such as diseases, pests, and pesticides (Guo et al., 2020). This is a key technology in the automatic mining of knowledge from very large Chinese agricultural texts and is the basis for downstream tasks such as building agricultural knowledge graphs and constructing agricultural intelligent question-and-answer (Q&A) systems (Drury and Roche, 2019; Guo et al., 2020). Therefore, accurate recognition of named entities in the field of kiwifruit plays an important role in ensuring the healthy development of the industry, plant protection, and convenience for agricultural workers.

Traditional NER methods can be divided into rule-based, dictionary-matching-based, and machine-learning-based approaches (Guo et al., 2020). Although each approach can achieve good results, they rely heavily on time- and energy-consuming pattern matching and feature engineering and have poor generalization ability. Through the application of deep learning in the field of NER and other fields (Chiu and Nichols, 2016; Bhatti et al., 2020b), researchers have developed various techniques for medical science (Zhao et al., 2019; Bhatti et al., 2021; Nawaz et al., 2021), cyber security (Li T et al., 2020), agriculture (Biswas and Sharan, 2021), social media (Aguilar et al., 2017) and environmental science (Bhatti et al., 2020a; Aamir et al., 2021; Galvan et al., 2022). In the field of Chinese NER (CNER), because sentences in Chinese texts are not naturally separated, unlike sentences in English, there is no obvious border symbol. Therefore, the first step in many original deep-learning-based CNER methods is to segment the text using word segmentation tools (Yang et al., 2016; He and Sun, 2017). With the development of research on CNER, many researches show that the character-based CNER model avoids segmentation errors and makes it more suitable than the word based model. (Jingzhou and Houfeng, 2008; Liu et al., 2010). However, in order to avoid the problem of segmentation errors, the character based CNER model cannot use Chinese word information. Recently many researchers have realized that word information will play a positive role in the correct recognition of Chinese entity boundaries. Therefore, lexicon-based CNER models have been widely used in recent years. For example, Zhang and Yang (2018) introduced the lattice long short-term memory model (Lattice-LSTM) based on a lexicon, allowing character-level and word-level information corresponding to the characters to be encoded simultaneously. Peng et al. (2020) proposed the Softlexicon method to integrate word information into the NER model by simply adjusting the character representation layer. The lexicon based model, with the help of the public lexicon, achieves better results than the purely

character based model (Peng et al., 2020). For example, when the lexicon based model recognizes the Chinese entity "长江大桥" (Yangtze River Bridge), words such as "长江" (Yangtze River), "大桥" (Bridge), and "长江大桥" (Yangtze River Bridge) in the lexicon can help eliminate the ambiguity of potentially related named entities in the context, such as the person name "江大桥" (Daqiao Jiang) (Zhang and Yang, 2018).

For CNER in the field of agriculture (CANER). The lexicon-based method makes good use of character information and word information, so using them to solve the CANER problem may be a theoretically feasible solution too. However, there is currently no open-source lexicon in the field of agriculture, and manual lexicon construction is labor-intensive. If the lexicon is built through automatic word segmentation, the existing word segmentation tools face the problem of word segmentation errors caused by insensitive word segmentation. For example, farm chemicals entities such as "速乐硼、辛硫磷乳油" (solubor, phoxim) and kiwifruit variety entities "中华猕猴桃、红心猕猴桃" (Actinidia chinensis Planch., red-fleshed kiwi), which exist in kiwifruit-related texts, have strong domain characteristics, and these will make the word segmentation tool insensitive in the form of out-of-vocabulary (OOV) words. Therefore, many CANER methods are still character-based models (Guo et al., 2020; Zhao et al., 2021; Guo et al., 2022), and the use of word information is hindered by word segmentation errors. As for the sequence coding layer of recently CANER model, bidirectional long short-term memory (BiLSTM) is still the mainstream deep learning method, which can memorize long-text sequence features in theory (Liu et al., 2020; Zhao et al., 2021). However, the contextual feature extraction ability of BiLSTM has the following limitations. First, with an increase in sentence length, the feature extraction ability of BiLSTM will decline (Li Y et al., 2020). Second, BiLSTM makes each character contribute equally to the task (Guo et al., 2020), but the contribution of different types of characters in agricultural texts to the task is certainly different. Third, the strong domain features of kiwifruit-related text, particularly farm chemical-related entities, disease-related entities, and pest-related entities, pose a challenge to the feature extraction ability of BiLSTM. In summary, deep learning-based methods for CANER in the field of kiwifruit diseases and pests face the following problems: The use of word information is hampered by OOV problem in the process of lexicon construction. And the contextual information capture capability of the sequence encoding layer needs to be further improved.

This research proposes a lexicon-based CANER model KIWINER on the basis of bidirectional long short term memory and conditional random field model (BiLSTM-CRF). The objectives of KIWINER are to take measures to solve the above problems in the end of the previous paragraph, that is, to integrate the word information containing domain features into the model, improve the model feature extraction ability, and ultimately provide support for the construction of the kiwifruit

Q&A system. Specifically, KIWINER improves the recognition quality through statistics-based new word detection, AttSoftlexicon, and PCAT. First, statistics-based new word detection is innovatively used to detect new words in kiwifruit-related text corpora, thereby improving the adaptability of word segmentation tools to kiwifruit-related texts and reducing the impact of word segmentation errors on the lexicon construction process; Second, through the AttSoftlexicon method proposed in this paper, based on Softlexicon (Peng et al., 2020) and CCNet (Huang et al., 2019), the character and word information in the lexicon are integrated into the model, and the position information of the character in the corresponding words can be fully utilized with the help of CCNet (Huang et al., 2019); Third, a novel module parallel connection criss-cross attention network (PCAT) is proposed to improve the contextual feature extraction ability of BiLSTM. PCAT assigns different weights to different characters according to their correlation and constructs a parallel structure through convolutional layers with different filter sizes to obtain richer semantic information. Additionally, this study collected publicly available textual information and constructed a kiwifruit NER dataset consisting of 17809 entities across six categories. Previous CANER methods based on machine learning, such as CRF (Li et al., 2017), rely on manual features or rules, which are time-consuming and unable to process a large number of complex agricultural texts (Guo et al., 2020). The CANER methods such as Att-BiLSTM-CRF (Zhao et al., 2021) use the deep learning method to reduce the work of designing feature extractors for each problem and solve the above problems. Compared with the popular CANER methods based on deep learning, our proposed KIWINER alleviates the OOV problem through new word detection, and makes full use of lexical information and agricultural features in addition to character information through AttSoftlexicon and PCAT, so the feature extraction ability of deep learning model is effectively improved. We also use KIWINER and five typical CNER models and two popular CANER models for comparative experiments, and the KIWINER model yields better performance.

The remainder of this paper is organized as follows. The materials used in this study and the methods proposed in this paper is discussed in detail in section 2. Section 2 also introduces the experimental parameters, dataset division, evaluation metrics, and the experimental environment. The experimental details and results are presented in Section 3. The discussion of this study is presented in section 4. Finally, the conclusions are presented in Section 5.

2 Materials and methods

The overall architecture of KIWINER, shown in Figure 1, indicates that the model contains six layers and uses BiLSTM-CRF as the basic framework. This section first introduces the

experimental materials. Then this section focuses on the implementation details of the new word detection layer, embedding layer, CCNet, and AttSoftlexicon and PCAT proposed in this paper. Details of the BiLSTM and CRF layers can be found in (Huang et al., 2015).

2.1 Materials

To solve the problem of the limited public NER dataset for CANER, a new kiwifruit-related annotated corpus, named KIWID, was collected and annotated under the guidance of plant protection experts from Northwest A&F University.

2.1.1 Corpus collection

To ensure the quality of data, this study collected public information on kiwifruit diseases and pests from the official websites of trusted research institutions and Baidu Encyclopedia. Preprocessing was applied to remove non-useful content, such as webpage tags, links, and special characters contained in the corpus. Finally, a corpus (Corpus A of kiwifruit) containing 61103 sentences for training character vectors and detecting new words was obtained.

2.1.2 Corpus tagging

We selected 12477 sentences from Corpus A to form Corpus B. Under the guidance of plant experts from Northwest A&F University, six types of kiwifruit-related entities were labeled, as shown in Table 1. Therefore, this study used the BMES (Ratinov and Roth, 2009) tagging scheme to tag Corpus B, where B, M, E, and S represent the beginning, middle, and end of an entity, and a single-word entity, respectively. To ensure annotation quality, the manual annotation method was adopted. Finally, the kiwifruit-related dataset KIWID containing 17809 entities was obtained, and the statistical information of KIWID is presented in the last column of Table 1.

2.1.3 Analysis of Corpus features

(1) Contains several specialized vocabulary terms.

Entities involved in agricultural diseases and pests such as farm chemicals entities, pest entities, plant disease entities, and varieties entities are annotated in the corpus, such as "二甲吗啉" (dimethomorph), "联苯菊酯" (bifenthrin), "介壳虫" (scale insect), and "斑点病" (scab). Such words usually do not appear in the built-in dictionaries of common word-segmentation tools and have strong domain characteristics. Therefore, most word segmentation tools have poor adaptability to these specialized terms, leading to a greater likelihood of word segmentation errors. If the word information in the lexicon constructed by automatic word segmentation is introduced into the CANER model, the accuracy of the model may be significantly affected by word segmentation errors.
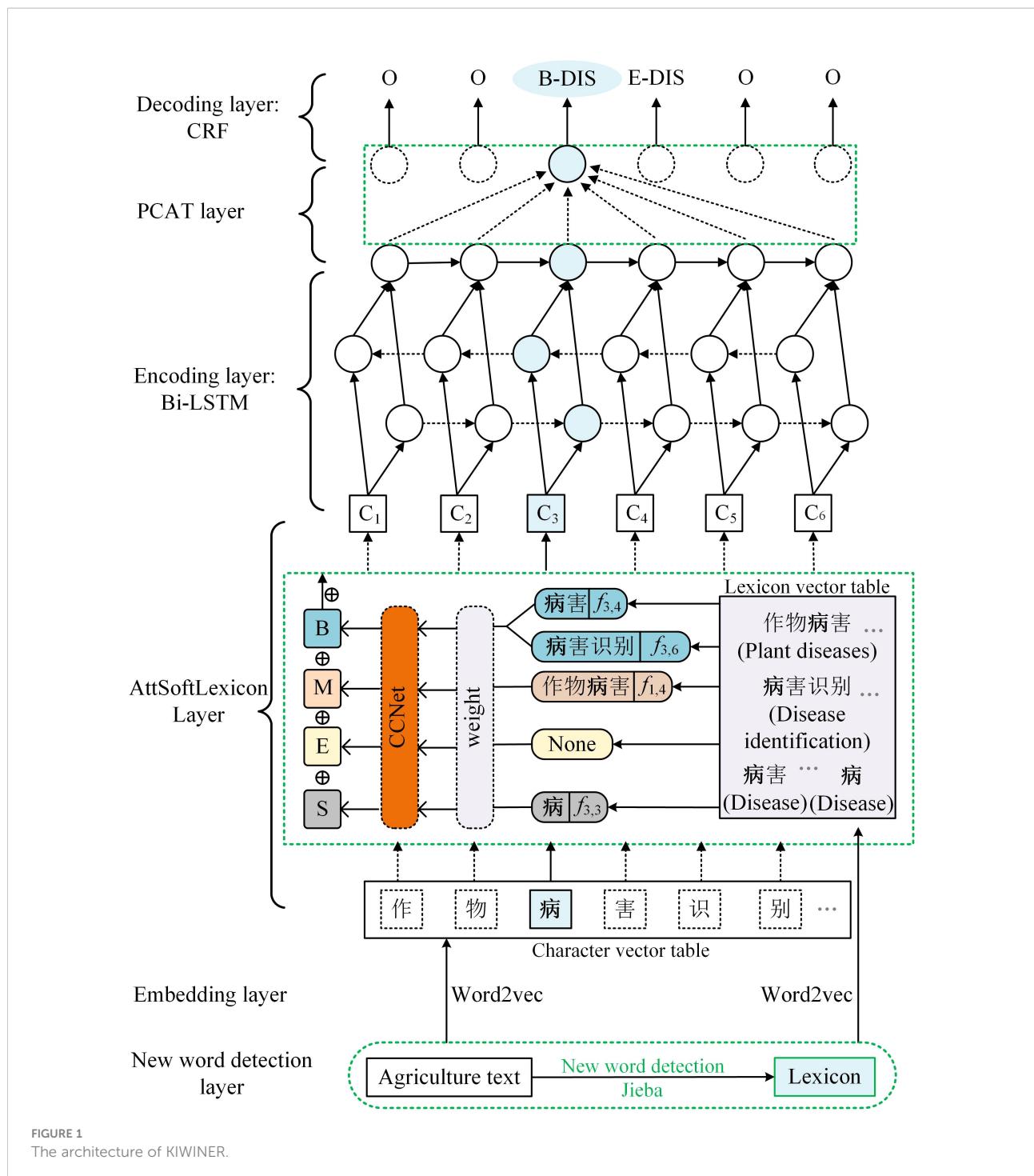
FIGURE 1
The architecture of KIWINER.

(2) Number of entities is unevenly distributed.

As shown in Table 1, there are differences in the number of different types of entities. The same problem exists not only in agriculture (Guo et al., 2022) but also in clinical medicine (Kong et al., 2021). The uneven distribution of the number of entities introduces challenges to the feature extraction ability of the CANER models.

(3) Entities nested within each other

Nested named entities are a common problem in the field of NER in the task of identifying kiwifruit-related entities. For example, there are two entities nested in "中华猕猴桃" (Actinidia chinensis Planch.), which are the location entity "中华" (China) and the plant entity "猕猴桃" (kiwifruit). First, this leads to errors in word segmentation. For example, Jieba's word

TABLE 1 Statistics of KIWID.

Category (Symbol)	Category definition	Examples	Numbers
Varieties (KIWI)	Names of different varieties of kiwifruit.	陇南猕猴桃 (Longnan kiwi)	3763
Disease (DIS)	Diseases of kiwifruit.	叶枯病 (Leaf blight)	561
Pest (PEST)	Pests of kiwifruit.	叶蝉 (Leaf cicada)	1247
Part (PART)	Diseases harming kiwifruit parts.	叶片 (Leaves)、枝干(branches)	5521
Farm chemical (MED)	Farm chemicals.	多菌灵 (Carbendazim)	907
Place (LOC)	Distribution area of kiwifruit	陕西 (Shaanxi)	5090

segmentation result of "中华猕猴桃" (Actinidia chinensis Planch.) is "中华 猕猴桃" (China kiwifruit). If the lexicon for the NER model contains incorrect word segmentation information, it provides misleading information for the identification of entity boundaries. Moreover, the phenomenon of nested entities also increases the difficulty of entity recognition and introduces challenges to the feature extraction ability of the model.

2.2 New word detection layer

New word detection can identify OOV words and add them to the built-in dictionary of the word segmentation tool, thus improving the effect of common word segmentation tools (Du et al., 2016). Currently, new word detection is either rule-based (Huiming et al., 2003), statistics-based (Jin and Tanaka-Ishii, 2006), or based on both rules and statistics (Zheng and Wen-Hua, 2002). Methods that rely entirely or partly on rules rely on a manually built rule base. Although the rule base is helpful in improving the effectiveness of new word detection, the construction process is complex and time-consuming, and domain transferability is poor. As a result, this study adopts a statistics-based new word detection method. Corpus A was first segmented into strings using the *N-gram* method, and the garbage strings were then filtered in turn according to the three statistics of word frequency (*WF*), mutual information (*MI*), and contextual entropy (*CE*) of the strings. Subsequently, a new word set was obtained. This new word set was then added to the built-in dictionary of Jieba to improve its applicability to kiwifruit-related texts. Finally, the kiwifruit lexicon was constructed through the word segmentation of Corpus B by Jieba. This section first introduces the methods related to new word detection, and then introduces the lexicon construction process.

2.2.1 *N-gram* Word segmentation

The basic idea of *N-gram* word segmentation is to use a fixed window of length *n* to segment the sentence. After segmentation, each string of size *N* is called a "gram." For example, the 2-gram segmentation result of the sentence "农业病害识别" (agricultural disease identification) is "农业/业病/病害/害识/识别" (nong ye/

ye bing/bing hai/hai shi/shi bie). Other examples are shown in Figure 2A.

2.2.2 Mutual information

The concept of *MI* originates from information theory and is commonly used to measure how consistently two patterns occur together in a corpus (Ye et al., 2013). The *MI* value is derived from the log-likelihood ratio of the joint probability of patterns *A* and *B* over the individual probabilities of patterns *A* and *B*, as shown in Equation (1). If Chinese strings w_1 and w_2 in the same dataset appear as a whole string w_{12}, the probability is $p(w_{12})$, and the probabilities of the two strings appearing alone are $p(w_1)$ and $p(w_2)$, respectively. The *MI* value was calculated using formula (2). The higher the *MI* value of the two strings, the more likely they are to be combined into meaningful words.

$$MI(x, y) = \log_2 \frac{p(x, y)}{p(x)p(y)}, \tag{1}$$

$$MI(w_1, w_2) = \log_2 \frac{p(w_{12})}{p(w_1)p(w_2)}. \tag{2}$$

2.2.3 Contextual entropy

CE is an external statistic proposed by (Huang and Powers, 2003), that can be used to measure the probability of whether a string is a meaningful word. It measures the randomness of the left and right adjacent characters of a string, that is, the left and right contextual entropies. Compared with a Chinese string with no practical meaning, a Chinese word with a practical meaning has a wider application scenario. Thus, the randomness of the set of left and right adjacent characters will be higher. Therefore, a higher *CE* value for a Chinese string indicates a greater probability that the string has a practical meaning. In the Chinese new word detection task, the *CE* accurately reflects the probability that a string is a meaningful word. The *CE* value was calculated using Equations (3) and (4):

$$E_l(w) = -\sum_{w_l \in S_l} P(w_l|w) \times \log_2 P(w_l|w), \tag{3}$$

$$E_r(w) = -\sum_{w_r \in S_r} P(w_r|w) \times \log_2 P(w_r|w), \tag{4}$$
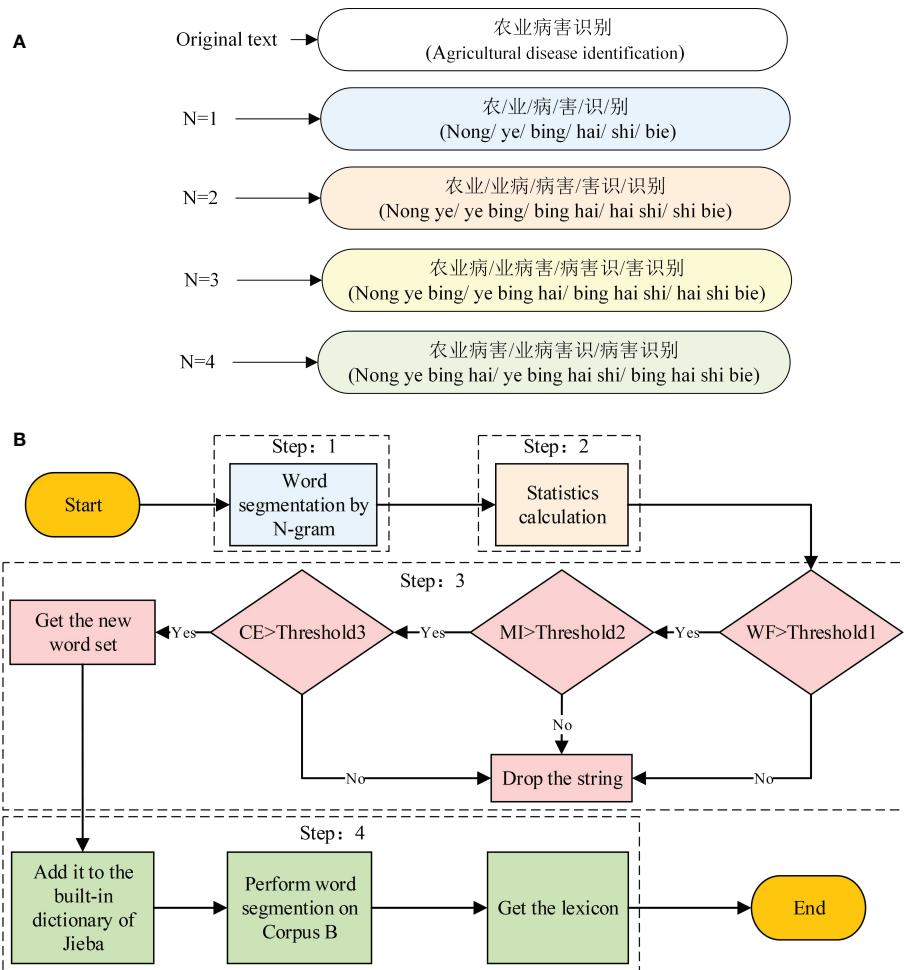
FIGURE 2
(A, B) Lexicon construction process.

where $p(w_l|w)$ represents the probability that the left-adjacent character of w is character w_l, $p(w_r|w)$ represents the probability that the right-adjacent character of w is character w_r, S_l represents all left-adjacent characters of w, and S_r represents all right-adjacent characters of w.

2.2.4 Lexicon construction

The lexicon construction process occurs in four steps, as illustrated in Figure 2B).

Step 1: Apply the *N-gram* word segmentation method to segment corpus A and obtain candidate strings with $N = 2, 3$, and 4.

Step 2: Calculate the statistics for each string. Compute the *WF*, *MI*, and *CE* values for each candidate string.

Step 3: Set the corresponding thresholds for *WF*, *MI*, and *CE*, named *Threshold1*, *Threshold2*, and *Threshold3*, respectively, and filter the candidate strings to obtain a new set of words.

To avoid the omission of low-frequency new words, we set the *WF* threshold to 5, *MI* threshold to 3.9, and *CE* threshold to 2.7.

Step 4: Add the new word set obtained in Step 3 to the built-in dictionary of Jieba and perform word segmentation on Corpus B to obtain the kiwifruit lexicon for NER.

2.3 Embedding Llayer

For a character-based CNER model, discrete text sequences are converted into low-dimensional densely distributed embedded representations, allowing the model to learn more semantic knowledge and improve its performance (Guo et al., 2020). As shown in Figure 1, to obtain a high-quality embedded representation and make good use of the information in the corpus, Word2vec-CBOW (Mikolov et al., 2013) was used to train Corpus A in character form and transform the resulting agricultural lexicon into vectors. The input sequence of length n

is $s=(c_1, c_2, c_3,\ldots\ldots,c_n) \in V_c$, where V_c is the word set (including characters), and each word is represented by a trained dense vector $x_i^c = e^c(c_i)$, where e^c denotes the word embedding lookup table.

2.4 CCNet

CCNet (Huang et al., 2019) is often used in semantic segmentation to aggregate contextual information from all pixels to obtain dense contextual information. This study considered the use of CCNet for text feature extraction. The overall structure of the CCNet is shown in Figure 3A.

Given a feature map $M \in R^{C \times W \times H}$, CCNet first generates two feature maps Q and K by applying two convolutional layers with a filter size of 1×1 on the feature map M. $\{Q, K\} \in R^{C' \times W \times H}$, where C' is the number of channels of Q and K, which is less than C for dimension reduction. Another convolutional layer with filters of size 1×1 is applied on M to generate $V \in R^{C \times W \times H}$. $Q_u \in R^{C'}$ is the vector for each position u in the spatial dimension of the feature map Q. And vector set $\Omega_u \in R^{(H+W-1) \times C'}$ is obtained by extracting feature vectors from K which are in the same row with position u. Then, CCNet can obtain $D \in R^{(H+W-1) \times W \times H}$, which represents the degree of correlation between features Q_u and $\Omega_{i,u}$ ($i=[1,\ldots,|\Omega_u|]$) by the affinity operation, which is defined as follows:

$$d_{i,u} = Q_u \Omega_{i,u}^T \qquad (5)$$

where $d_{i,u} \in D$. Feature map A is then obtained by applying a softmax layer on D over the channel dimension. CCNet can also obtain vector $V_u \in R^C$ and set $\theta_u \in R^{(H+W-1) \times C}$. The set θ_u is a collection of feature vectors in V that are in the same row as position u. Finally, the contextual information is collected by the aggregation operation:

$$M_u' = \sum_{i \in |\theta_u|} A_{i,u} \theta_{i,u} + M_u \qquad (6)$$

where $M_{u'}$ is a feature vector in the output feature maps $M' \in R^{C \times W \times H}$ at position u, and $A_{i,u}$ is a scalar value at channel i and position u in A. Contextual information is added to local feature M to enhance the local features and augment the pixel-wise representation.

2.5 Criss-cross attention based Softlexicon layer

One of the tasks of CANER is to recognize the boundaries of agricultural entities, and word segmentation information provides good guidance for identifying entity boundaries. However, CANER is affected by the strong domain characteristics of agricultural texts and the uneven distribution of entity categories (Guo et al., 2020). Adding more pre-training information will help the model learn more agricultural characteristics, thus reducing the impact of the aforementioned problems. Therefore, this paper proposes an AttSoftlexicon based on Softlexicon (Peng et al., 2020) and CCNet (Huang et al., 2019),
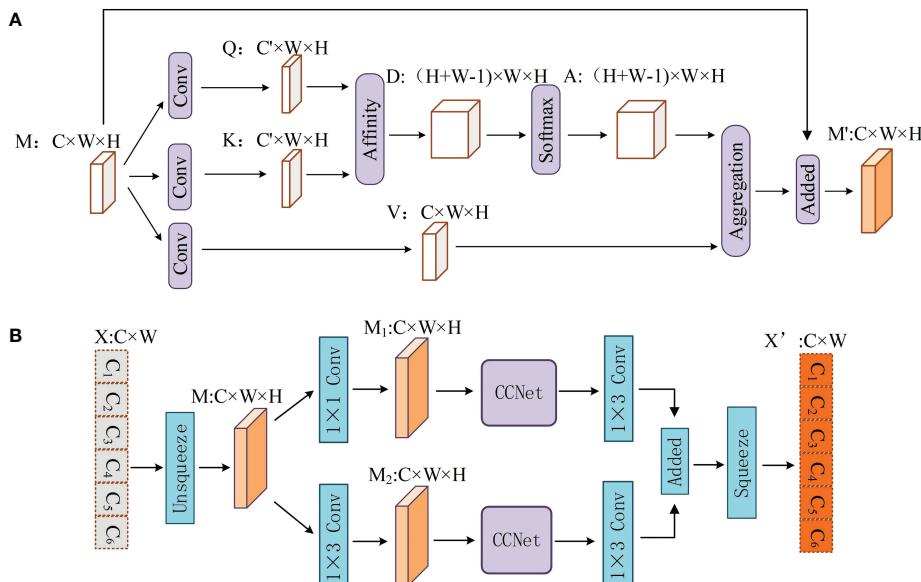
FIGURE 3
(A, B) Structure of CCNet and PCAT.

and integrates the word information in the lexicon into character representation, which helps the model to learn more kiwifruit text features.

Assume that the input sequence is $s=\{c_1, c_2,\ldots, c_n\}$, and $w_{i,j}$ denotes its subsequence $\{c_i, c_{i+1},\ldots, c_j\}$. The first step is lexicon matching. Each character is matched from a lexicon to all words containing the character. According to the position of each character c_i in the different matched words (beginning, middle, end, or one-character word), the words matched by a character were divided into four-word sets $B(c_i)$, $M(c_i)$, $E(c_i)$, and $S(c_i)$. The set construction method is shown in formula (7)-(10).

$$B(c_i) = \left\{ w_{i,k}, \forall\ w_{i,k} \in L, i < k \leq n \right\}, \tag{7}$$

$$M(c_i) = \left\{ w_{j,k}, \forall\ w_{j,k} \in L, 1 \leq j < i < k \leq n \right\}, \tag{8}$$

$$E(c_i) = \left\{ w_{j,i}, \forall\ w_{j,i} \in L, 1 \leq j < i \right\}, \tag{9}$$

$$S(c_i) = \left\{ c_i, \exists\ c_i \in L \right\}. \tag{10}$$

As shown in formula (7)-(10), L denotes the lexicon, and w represents the words matched in the lexicon. If a word set of characters is empty, it is represented as *{None}*. Taking the input sequence "植物病害" (plant disease) as an example, the character "物" (matter) is matched with the pre-constructed lexicon, and the two words "植物病害" (plant disease) and "植物" (plant) are matched, and the four word sets corresponding to the character "物" (matter) are formed: $B=\{$"None"$\}$, $M=\{$"植物病害"$\}$, $E=\{$"植物"$\}$, $S=\{$"None"$\}$. *The character "病" (disease) is matched with the pre-constructed lexicon, and the two words "病害" (disease and pest) and "病" (disease) are matched, and the four-word sets corresponding to the character "病" (disease) are formed:* $B=\{$"病害"$\}$, $M=\{$"None"$\}$, $E=\{$"None"$\}$, $S=\{$"病"$\}$, *as shown in* Figure 4. *To integrate the word set information matched to each character into the corresponding character representation, the statistics-based static weighting method in Softlexicon* (Peng et al., 2020) *was used, where the frequency reflects the importance of the word.*

The weighting method is given by formulae (11) and (12), where $z(w)$ is the frequency with which a lexicon word w occurs in the statistical data and e^w is the word embedding lookup table. The weighted representation of word set S is obtained as follows:

$$v^s(S) = \frac{4}{Z} \sum_{w \in S} z(w) e^w(w), \tag{11}$$

Where:

$$Z = \sum_{w \in B \cup M \cup E \cup S} z(w). \tag{12}$$

In the last step, the original Softlexicon (Peng et al., 2020) combines the representations of four-word sets into the fix-dimensional feature and adds it to the representation of each character, as shown in formulae (13) and (14).

$$V = [v^s(B); v^s(M); v^s(E); v^s(S)], \tag{13}$$

$$x^c \leftarrow [x^c; V]. \tag{14}$$

The original Softlexicon (Peng et al., 2020) designed four-word sets to take advantage of these four types of positional information. However, it only weighs the words in each word set according to the word frequency and does not distinguish the importance of different word sets. This does not allow the model to distinguish the four positions of the characters in the matched words.

CCNet (Huang et al., 2019) showed a strong contextual relationship extraction ability in the semantic segmentation task. Therefore, to make full use of these four types of position information, this study uses CCNet to learn the weights for different word sets, as shown in the formula (15). First, CCNet processes the representation of these four sets and automatically assigns weights to them based on the relationship between them. It is then transformed into a vector of 1×4 through q. Finally, the weight vector a_i ($i \in [1,4]$) with a value range of (0, 1) is obtained through the sigmoid function. a_i is a weight matrix of dimensions 1×4, where the four values represent the importance of the four word sets. As shown in formula (16), the four-word set representations are weighted and merged into the character representation.

$$a_i = sigmoid(qCCNet(V)) \tag{15}$$

$$x^c = [x^c; a_1 v^s(B); a_2 v^s(M); a_3 v^s(E); a_4 v^s(S)]. \tag{16}$$

2.6 Parallel connection Criss-cross attention network

The sequence features extracted by BiLSTM may have a few limits. First, with an increase in sentence length, the feature extraction ability of BiLSTM declines (Huang et al., 2019). In addition, LSTM has been shown to have weaker feature extraction ability than attention mechanism models, such as transformers, when dealing with longer sequence texts (Li et al., 2020b). Second, BiLSTM makes each character contribute equally to the task. In other words, BiLSTM is not good at assigning more weight to some important characters in the text sequence, which is very important for NER. In addition, the strong domain features of kiwifruit-related texts mentioned in Section 2.1.3 also pose challenges to the feature extraction ability of the BiLSTM. In short, the feature extraction ability of BiLSTM must be further improved when solving the problem of kiwifruit-named entity recognition. Therefore, a novel module, parallel connection criss-cross attention network (PCAT), is proposed to mitigate the impact of the above limits with the help of CCNet (Huang et al., 2019). The overall structure of the PCAT is shown in Figure 3B.
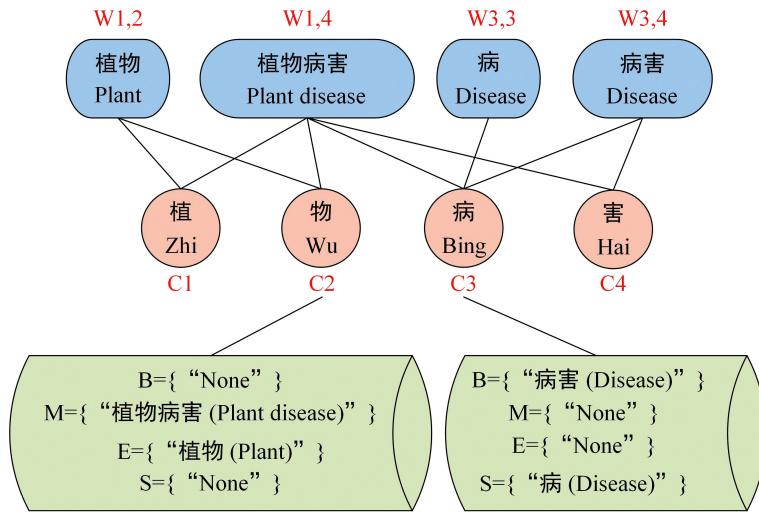
FIGURE 4
Example of lexicon matching.

After the agricultural sentence is processed by BiLSTM, a feature map $X \in R^{C \times W}$ is obtained (C represents the dimension of BiLSTM and W represents the length of the sentence). In this work, agricultural sentences are regarded as pictures with a channel number of C and a size of $W \times 1$. Therefore, PACT first transforms X into a feature map $M \in R^{C \times W \times H}$ (the value of H is 1) through an unsqueeze operation. Each pixel in the feature map M represents a character in the agriculture text.

To obtain richer semantic information. The PCAT uses two different convolutional layers with filter sizes of 1×1 and 1×3 on M to generate two feature maps, M_1 and M_2. M_1 and M_2 are put into the CCNet for processing. To learn more complex features, PCAT applies two convolutional layers with filter sizes of 1×3 to M_1 and M_2. Finally, M_1 and M_2 are added, and the output vector of the PCAT $X' \in R^{C \times W}$ is obtained through a squeeze operation.

Using CCNet to calculate the connection between each character, PCAT can assign different weights to different characters to give more attention to key characters. In addition, PCAT can solve the problem of long-distance dependency because it can calculate the degree of association between words in each position and other words that are not affected by distance. Through a parallel structure and convolutional layer, PCAT can obtain richer features from agricultural texts.

2.7 Evaluation indicators and experimental environment

2.7.1 Parameter setting

In our proposed model, both the character vector dimension and word vector dimension were set to 50. In the feature encoding

layer, the hidden size of both the forward and backward LSTM was set to 300, and to mitigate overfitting, the dropout rate was set to 0.5. For the model training, the batch size was set to 16. Furthermore, the model was trained using stochastic gradient descent with an initial learning rate of 0.0015, and the learning rate decay was set to 0.05. The hyper-parameter configuration of the model is listed in Table 2. All experiments were conducted under the conditions listed in Table 3.

2.7.2 Dataset division

For dataset division, four datasets were involved in the experiment, namely KIWID, BOSON, ClueNER, and People's Daily. We obtained the public data according to Table 2 in study (Liu et al., 2022). This study randomly divided KIWID, BOSON, and ClueNER into training, validation, and test sets according to a ratio of 8:1:1, respectively [refer to Zhang et al. (2021)]. Division of People's Daily reference https://github.com/zjy-ucas/ChineseNER. The pre-training corpus used in the KIWID-related experiments was the kiwifruit pre-training corpus constructed in this study. The pre-training corpus used in public dataset-related experiments is derived from Lattice-LSTM (Zhang and Yang, 2018), which is pre-trained using Word2vec (Peng et al., 2020) over automatically segmented Chinese Giga-Word. The number of character vectors in the public pre-training corpus is 5.7k, and the number of words in the lexicon is 704.4k.

2.7.3 Evaluation indicators

Precision (P), recall (R) and F_1-score (F_1) were used to evaluate the performances of the different models, as shown in Equations (17)-(19).

TABLE 2 Hyper-parameter value.

Parameters	Value	Parameters	Value
character embedding dim	50	learning rate decay	0.05
batchsize	16	LSTM hidden	300
learning rate	0.0015	dropout rate	0.5

TABLE 3 Experimental environment.

Project	Environment	Project	Environment
Operating system	Windows 10(x64)	Hard disk	1T
CPU	i7-10700F@2.90GHz	Python version	3.6.5
GPU	NVIDIA TITANRTX (24GB)	Pytorch version	1.8.1
Memory	64GB	–	–

$$P = \frac{\text{True positives}}{\text{Predicted as positives}} = \frac{T_P}{T_P + F_P}, \quad (17)$$

$$R = \frac{\text{True positives}}{\text{Actual positives}} = \frac{T_P}{T_P + F_N}, \quad (18)$$

$$F_1 = \frac{2PR}{P + R}. \quad (19)$$

True positives (T_P) refer to the number of correctly recognized positive samples among all positive samples, whereas false positives (F_P) denote the number of negative samples incorrectly recognized as positive samples. False negatives (F_N) are positive samples incorrectly recognized as negative samples. Among all the positive samples, the more that are predicted correctly, the higher the P value. A higher number of positive samples predicted in the testing set yielded a higher R value. F_1 is the harmonic average of P and R, providing an evaluation of the comprehensive ability of the model.

3 Results

3.1 Experiments on KIWID

In this section, some typical NER models such as BiLSTM (Huang et al., 2015), TENER (Yan et al., 2019), LR-CNN (Gui et al., 2019a), LGN (Gui et al., 2019b) and Softlexicon-LSTM (Peng et al., 2020) are considered comparable models. In addition, this section also uses the previous CANER findings JMCA-ADP (Guo et al., 2020) and Att-BiLSTM-CRF (Zhao et al., 2021) as comparison models. Like KIWINER, LR-CNN, LGN and Softlexicon-LSTM are also lexicon-based models. The lexicon used in the experiments in this section are the Kiwifruit lexicon constructed in this study.

The experimental results for KIWID are shown in Table 4. It could be observed that the model proposed in this study

outperformed other models, and the F_1 of this model is at least 0.47 higher than other models, which illustrates the effectiveness of it recognizing kiwifruit-related entities. The performance of our model is significantly improved compared to the baseline model BiLSTM-CRF. This is due to the fact that KIWINER makes full use of kiwifruit lexical information with the help of AttSoftlexicon, and obtains deeper semantic features with the help of PCAT. Compared with CANER models Att-BiLSTM-CRF and JMCA-ADP, KIWNER has achieved obvious improvement, which further verifies the effectiveness of KIWINER. The lexicon-based models LR-CNN, LGN, Softlexicon-LSTM and KIWINER have clear advantages over the rest of the character-based models, illustrating the effectiveness of constructing a kiwifruit-related lexicon and incorporating lexical information into the model.

3.2 Experiments on public datasets

To verify the generalization of KIWINER, three public datasets were selected: Boson, ClueNER, and People's Daily. The experimental results are listed in Table 5.

The KIWINER model achieved the best F_1 of the three datasets, which were for Boston, ClueNER, and People's Daily 85.13%, 80.52%, and 92.82%, respectively. The experimental results show that KIWINER not only has performance advantages on the KIWID corpus, but also has a certain generalization in other fields.

3.3 Ablation experiments

3.3.1 Effectiveness of new word detection layer

In the new word detection layer of KIWINER, the adaptability of Jieba to kiwifruit-related texts was enhanced by new word detection and then a lexicon was constructed by word

TABLE 4 Results of each model on KIWID.

Model	P	R	F1
BiLSTM-CRF	84.42	84.54	84.48
Att-BiLSTM-CRF	82.85	88.99	85.81
JMCA-ADP	84.90	90.47	87.59
TENER	86.40	90.19	88.25
LR-CNN	87.08	89.90	88.47
LGN	86.81	89.63	88.19
Softlexicon-LSTM	87.18	89.27	88.21
KIWINER (our)	88.21	90.31	88.94

segmentation of kiwifruit-related texts. To verify the effectiveness of this lexicon construction method, this section used several commonly used Chinese automatic word segmentation tools (Pkuseg, Thulac, HanLP, Jieba, and Snownlp) to automatically separate the kiwifruit-related texts collected in this study to construct lexicons and apply them to KIWINER for experiments. Experiments were performed using KIWID. The experimental results are shown in Figure 5A.

The method of constructing the lexicon with the aid of new word detection and Jieba achieves the highest P, R, and F_1, and improves over other methods. This shows that new word detection effectively reduces the negative impact of word segmentation errors on CANER during lexicon construction.

3.3.2 Effectiveness of AttSoftlexicon layer

To verify the effectiveness of the AttSoftlexicon, it was replaced in KIWINER by Softlexicon (Peng et al., 2020), and a comparative experiment was conducted. The experiment used the F_1 as the evaluation metric, and the experimental results are shown in Figure 5B. The KIWINER model achieved the best F_1 for the four datasets. This shows that by assigning different weights to different word set representations, the AttSoftlexicon can help the model to make full use of the position information of characters in its matched words, thus making more full use of lexicon information than Softlexicon.

3.3.3 Effectiveness of PCAT layer

To verify the applicability of the PCAT module for different sequence encoding models, experiments were performed using transformer and GRU instead of BiLSTM. And comparative experiments were carried out with or without the PCAT module in the model. The experiment was divided into three groups, and the results are presented in Table 6.

The effect of each sequence coding model in the table improved after the introduction of PCAT, indicating the effectiveness and universality of PCAT. The model based on BiLSTM achieved the best effect, which shows the rationality of KIWINER using BiLSTM to encode character sequences.

4 Discussion

4.1 Comparison of experiments with different variants

To verify the rationality of the PCAT module, several variants of it were designed, and the variant was used to replace the PCAT in KIWINER for experiments on Boson, ClueNER, KIWID, and People's Daily. Variants A and B increased and decreased the depth of the PCAT, respectively. Variants C and D break the parallel connection structure of PCAT. The different variant structures of the PCAT are shown

TABLE 5 Results for each model on public datasets.

Model	Boson			ClueNER			People's Daily		
	P	R	F1	P	R	F1	P	R	F1
LSTM	81.78	72.50	76.86	76.80	71.28	73.94	85.96	82.09	83.98
Att-BiLSTM-CRF	79.93	76.67	78.27	74.73	73.62	74.17	86.28	85.05	85.66
JMCA-ADP	80.10	77.66	78.86	75.82	76.58	76.20	87.96	86.93	87.44
TENER	79.45	81.51	80.47	74.34	77.08	75.68	90.36	90.07	90.22
LR-CNN	84.40	82.04	83.20	80.09	78.47	79.27	91.13	90.74	90.93
LGN	82.16	79.16	80.63	77.01	73.95	75.45	90.75	89.52	90.13
Softlexicon-LSTM	85.75	80.67	83.13	80.50	79.11	79.80	92.31	90.43	91.36
KIWINER	86.96	83.37	85.13	81.05	80.01	80.52	93.23	92.42	92.82

FIGURE 5
(A, B) Results for each lexicon construction method & Effectiveness of AttSoftlexicon.

in Figure 6. In addition, many researchers use the self-attention mechanism (Self-Att) to improve the feature extraction ability of the sequence encoding layer. In the field of CANER, Guo et al. (2020) introduced a self-attention module after the BiLSTM model to improve the feature extraction ability of sequence coding layer. Therefore, this section refers to the study by Guo et al. (2020) and uses Self-Att instead of PCAT for experiments. Attention unit and head number of Self-Att is 600 and 8. The experimental results are listed in Table 7.

Compared with Variants A and B, the PCAT achieved better results, indicating that the depth design of the PCAT is reasonable. Compared with variables C and D, PCAT achieves better results, which shows that a parallel structure can effectively improve the feature extraction ability of the model and help the model obtain richer semantic information. PCAT achieves better results than Self-Att (Guo et al., 2020), which indicates that PCAT is more conducive to improving the model feature extraction capability than the commonly used module Self-Att. PCAT constructs a parallel structure with the help of two different convolutional layers, which allows the model to simultaneously process semantic information from two different perspectives. At the same time, with the help of CCNet, which has good long distance context semantic aggregation capability (Huang et al., 2019), the information can be processed again, and different weights can be given according to different information relationships. Therefore, PCAT can help the model make full use of the feature information input into the model.

4.2 Comparative analysis with the previous CANER findings

This section discusses the recognition effects of KIWNER and the previous CANER studies Att-BiLSTM-CRF (Zhao et al., 2021) and JMCA-ADP (Guo et al., 2020) on each category of the kiwifruit dataset KIWID. BiLSTM-CRF (Huang et al., 2015), the baseline model of the above models, also participated in the experiments. The experimental results are shown in Table 8, where F_1 is chosen as the evaluation metric, and the last column of the table is the running time of each model.

It can be clearly seen from the table that KIWNER has achieved the best results in each category, especially in disease, pest, pesticide, which contain strong domain features. Although Att-BiLSTM-CRF and JMCA-ADP have made efforts to integrate agricultural features into the model, KIWINER can obtain more agricultural features by using word information with the help of Attsoftlexicon and new word detection. In addition, PCAT can help the model to further make full use of these agricultural features. The category of location related entities usually contain boundary characters, such as "县" (county), "镇" (town), "村" (village), etc., and the category of part related entities have limited diversity and many repeated words, which leads to the recognition difficulty of the above two categories being relatively low. Therefore, KIWINER did not significantly improve the recognition effect of LOC and PART. From the last row of the table, we can see that KIWINER takes more time than other models, which is a disadvantage of

TABLE 6 Application effect of PCAT.

Group	Model	F_1	Model	F_1
1	AttSoftlexicon-Transformer-CRF	84.01	AttSoftlexicon-Transformer-PCAT-CRF	85.11
2	AttSoftlexicon-BiGRU-CRF	87.68	AttSoftlexicon-BiGRU-PCAT-CRF	88.85
3	AttSoftlexicon-BiLSTM-CRF	87.17	AttSoftlexicon-BiLSTM-PCAT-CRF(KIWINER)	88.94
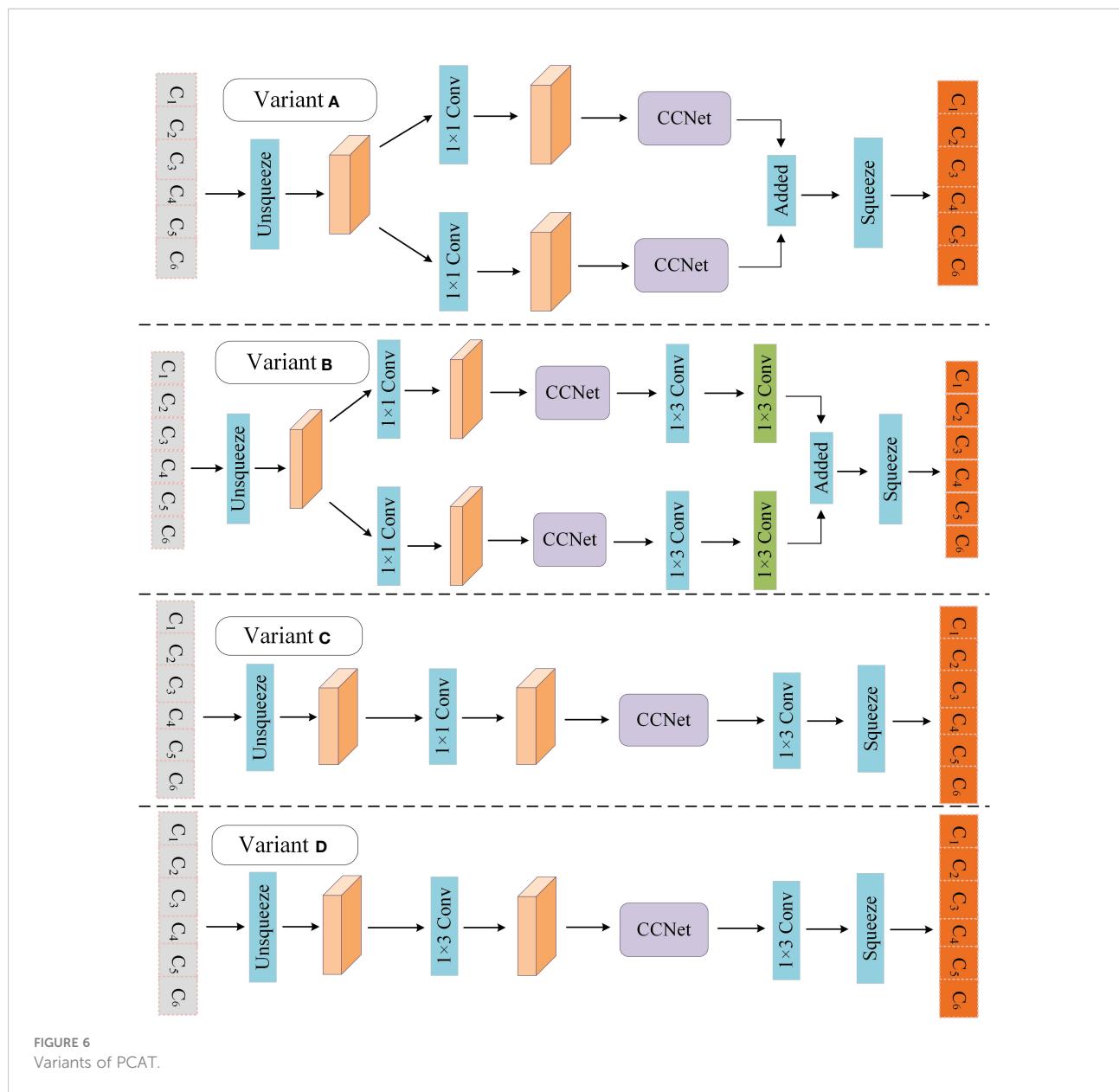
FIGURE 6
Variants of PCAT.

KIWINER. KIWNER incorporates lexical information, so it will spend an extra part of time on processing lexical information compared with the character based model. Research on faster character and word matching methods and more efficient sequence encoding modules can be helpful to overcome this shortcoming.

In KIWNER, AttSoftlexicon module and PCAT module both adopt the CCNet model from semantic segmentation,

TABLE 7 Results for several variants of PCAT.

Module	Boson	ClueNER	KIWID	People's Daily
Self-Att	83.35	79.73	88.12	91.71
Variant A	84.03	80.26	88.41	92.34
Variant B	84.24	80.41	88.62	92.58
Variant C	84.07	80.32	88.22	92.69
Variant D	84.61	80.34	88.83	92.56
PCAT	**84.95**	**80.95**	**88.94**	**92.82**

The values in bold represent the maximum value in the same column.

TABLE 8 Entity categories study.

Entity type	BiLSTM	Att-BiLSTM-CRF	JMCA-ADP	KIWINER
KIWI	83.70	80.00	85.90	87.06
DIS	79.17	78.43	81.63	87.50
PEST	77.00	85.71	86.96	89.76
LOC	81.90	82.64	83.69	84.19
PART	94.44	94.23	94.43	96.80
MED	61.64	70.45	73.33	75.28
All category	84.48	85.81	87.59	88.94
Time(s/epoch)	139.14	149.20	144.13	163.87

and have achieved good results through experimental verification. With the help of new word detection and AttSoftlexicon, KIWINER incorporate the word information containing domain features into the model. And KIWINER has achieved significant improvement compared with previous which are character-based models. This shows that when solving problems with strong domain features such as CANER, it is a good solution to find a method to integrate more domain features into the model. In addition, the effectiveness of PCAT also shows the importance of making full use of these features.

5 Conclusion

To address the lack of an annotation dataset for agricultural named entity recognition in the kiwifruit field, a kiwi-annotated NER corpus KIWID, which contains six categories and 17089 entities was constructed in this study. According to the characteristics of kiwifruit-related texts, a new CANER model, KIWINER, was proposed by statistics-based new word detection and the novel module AttSoftlexicon, PCAT. To alleviate the word segmentation insensitivity caused by the strong specialization of kiwifruit-related texts, statistics-based new word detection was used to enrich the built-in vocabulary of Jieba and improve its applicability to kiwifruit texts to construct the kiwifruit lexicon. Inspired by the CCNet module in the field of semantic segmentation, the AttSoftlexicon was proposed to help the model make efficient use of lexicon information. In addition, this study proposes a PCAT module to improve the feature extraction ability of the sequence coding layer BiLSTM. The experimental results with the comparative models show that our proposed model can effectively improve CANER performance, particularly for difficult-to-recognize categories such as diseases, pests, and farm chemicals.

Moreover, our research can provide reference for developing new deep learning methods for named entity recognition of international texts. Theoretically, our construction method of Attsoftlexicon is also applicable for the named entity recognition of the texts of other similar languages, such as Japanese, Korean etc., which are unnaturally partitioned just like Chinese. In addition, our proposed PCAT module is used to improve the

sequence encoding ability of deep learning model essentially. So, applying our proposed PCAT module for the named entity recognition of other language is also theoretically feasible. Therefore, KIWINER can also be used to explore CNER tasks in other crops or other fields with domain features. In the future, we will study how to improve the time efficiency of KIWINER and use it in the construction of kiwifruit Q&A system.

Data availability statement

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

Author contributions

Conceptualization, LZ. Methodology, LZ. Validation, LZ. Formal analysis, LZ and XN. Investigation, LZ, XN, HZ, VG, CR and DN. Data curation, LZ, MZ and MG. Writing—original draft preparation, LZ. Writing—review and editing LZ and DN. Visualization, LZ and XN. Supervision, HZ and DN. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

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

Aamir, M., Li, Z., Bazai, S., Wagan, R. A., Bhatti, U. A., Nizamani, M. M., et al. (2021). Spatiotemporal change of air-quality patterns in hubei province–a pre-to post-COVID-19 analysis using path analysis and regression. *Atmosphere* 12 (10), 1338. doi: 10.3390/atmos12101338

Aguilar, G., Maharjan, S., Monroy, A. P. L., and Solorio, T. (2017). "A multi-task approach for named entity recognition in social media data," in *Proceedings of the 3rd Workshop on Noisy User-generated Text*, , Copenhagen, Denmark7 September. 148–153. doi: 10.18653/v1/W17-4419

Bhatti, U. A., Yuan, L., Yu, Z., Nawaz, S. A., Mehmood, A., Bhatti, M. A., et al. (2021). Predictive data modeling using sp-kNN for risk factor evaluation in urban demographical healthcare data. *J. Med. Imaging Health Inf.* 11 (1), 7–14. doi: 10.1166/jmihi.2021.3313

Bhatti, U. A., Yu, Z., Li, J., Nawaz, S. A., Mehmood, A., Zhang, K., et al. (2020a). Hybrid watermarking algorithm using clifford algebra with Arnold scrambling and chaotic encryption. *IEEE Access* 8, 76386–76398. doi: 10.1109/ACCESS.2020.2988298

Bhatti, U. A., Yu, Z., Yuan, L., Zeeshan, Z., Nawaz, S. A., Bhatti, M., et al. (2020b). Geometric algebra applications in geospatial artificial intelligence and remote sensing image processing. *IEEE Access* 8, 155783–155796. doi: 10.1109/ACCESS.2020.3018544

Biswas, P., and Sharan, A. (2021). A noble approach for recognition and classification of agricultural named entities using Word2Vec. *Int. J. Adv. Stud. Comput. Sci. Eng.* 9 (12), 1–8.

Chiu, J., and Nichols, E. (2016). Named entity recognition with bidirectional LSTM-CNNs. *Trans. Assoc. Comput. Linguist.* 4, 357–370. doi: 10.1162/tacl_a_00104

Drury, B. M., and Roche, M. (2019). A survey of the applications of text mining for agriculture. *Comput. Electron. Agric.* 163, 104864. doi: 10.1016/j.compag.2019.104864

Du, L., Li, X., Yu, G., Liu, C., and Liu, R. (2016). New word detection based on an improved PMI algorithm for enhancing segmentation system. *Acta Scientiarum Naturalium Universitatis Pekinensis.* 52 (1), 35–40. doi: 10.13209/j.0479-8023.2016.024

Galvan, L. C., Bhatti, U. A., Campo, C. C., and Trujillo, R. S. (2022). The nexus between CO2 emission, economic growth, trade openness: Evidences from middle-income trap countries. *Front. Environ. Sci.* 10. doi: 10.3389/fenvs.2022.938776

Gui, T., Ma, R., Zhang, Q., Zhao, L., and Huang, X. (2019a). "CNN-Based Chinese NER with lexicon rethinking," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, Macao, China, 4982–4988. doi: 10.24963/ijcai.2019/692

Gui, T., Zou, Y., Peng, M., Fu, J., Wei, Z., and Huang, X. (2019b). "A lexicon-based graph neural network for Chinese NER," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Hongkong, China, 1040–1050. doi: 10.18653/v1/D19-1096

Guo, X., Lu, S., Tang, Z., Bai, Z., Diao, L., Zhou, H., et al. (2022). CG-ANER: Enhanced contextual embeddings and glyph features-based agricultural named entity recognition. *Comput. Electron. Agric.* 194, 106776. doi: 10.1016/j.compag.2022.106776

Guo, X., Zhou, H., Su, J., Hao, X., and Li, L. (2020). Chinese Agricultural diseases and pests named entity recognition with multi-scale local context features and self-attention mechanism. *Comput. Electron. Agric.* 179, 105830. doi: 10.1016/j.compag.2020.105830

He, H., and Sun, X. (2017). "A unified model for cross-domain and semi-supervised named entity recognition in Chinese social media," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, USA, 3216–3222. doi: 10.1609/aaai.v31i1.10977

Huang, J. H., and Powers, D. (2003). "Chinese Word segmentation based on contextual entropy," in *Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation*, (Sentosa, Singapore: COLIPS PUBLICATIONS) 152–158.

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., and Liu, W. (2019). "CCNet: Criss-cross attention for semantic segmentation," in *Proceedings of the International Conference on Computer Vision*, Seoul, South Korea, 603–612.

Huang, Z., Wei, X., and Kai, Y. (2015). Bidirectional LSTM-CRF models for sequence tagging. *arXiv* [Preprint]. doi: 10.48550/arXiv.1508.01991

Huiming, D., Xiaojing, B., Baobao, C., and Shiwen, Y. (2003). "Chinese Word segmentation at peking university," in *Proceedings of the Second Workshop on Chinese Language Processing*, Sapporo, Japan, 152–155.

Jiang, Z., and Zong, C. H. (2020). A comprehensive understanding with the importance of popular science knowledge to the kiwifruit quality improvement. *China Fruits.* 01, 1–8. doi: 10.16626/j.cnki.issn1000-8047.2020.01.001

Jingzhou, H., and Houfeng, W. (2008). "Chinese Named entity recognition and word segmentation based on character," in *Proceedings of the Third International Joint Conference on Natural Language Processing*, Hyderabad, India, 128–132.

Jin, Z., and Tanaka-Ishii, K. (2006). "Unsupervised segmentation of Chinese text by use of branching entropy," in *Proceedings of the 21st International Conference on Computational Linguistics*, Sydney, Australia, 428–435.

Kong, J., Zhang, L., Jiang, M., and Liu, T. (2021). Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition. *J. Biomed. Inform.* 116, 103737. doi: 10.1016/j.jbi.2021.103737

Li, Y., Du, G., Xiang, Y., Li, S., and Chen, H. (2020). Towards Chinese clinical named entity recognition by dynamic embedding using domain-specific knowledge. *J. Biomed. Inform.* 106, 103435. doi: 10.1016/j.jbi.2020.103435

Li, T., Hu, Y., Ju, A., and Hu, Z. (2020). Adversarial active learning for named entity recognition in cybersecurity. *Comput. Mater. Continua.* 66, 407–420. doi: 10.32604/CMC.2020.012023

Liu, J., Gao, L., Guo, S., Ding, R., and Thiruvady, D. (2020). A hybrid deep-learning approach for complex biochemical named entity recognition. *Knowledge-based Systems.* 221, 106958. doi: 10.1016/j.knosys.2021.106958

Liu, P., Guo, Y., Wang, F., and Li, G. (2022). Chinese Named entity recognition: The state of the art. *Neurocomputing* 473, 37–53. doi: 10.1016/j.neucom.2021.10.101

Liu, Z., Zhu, C., and Zhao, T. (2010). "Chinese Named entity recognition with a sequence labeling approach: based on characters, or based on words?," in *Proceedings of the 6th International Conference on Intelligent Computing*, Changsha, China, 634–640. doi: 10.1007/978-3-642-14932-0_78

Li, X., Wei, X., Jia, L., Chen, X., Liu, L., and Zhang, Y. (2017). Recognition of crops, diseases and pesticides named entities in Chinese based on conditional random fields. *Trans. Chin. Soc. Agric. Machinery* 48 (s1), 178–185. doi: 10.6041/j.issn1000-1298.2017.S0.029

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv* [Preprint]. doi: 10.48550/arXiv.1301.3781

Nawaz, S. A., Li, J., Bhatti, U. A., Bazai, S. U., Zafar, A., Bhatti, M. A., et al. (2021). A hybrid approach to forecast the COVID-19 epidemic trend. *PloS One* 16 (10), e0256971. doi: 10.1371/journal.pone.0256971

Peng, M., Ma, R., Zhang, Q., and Huang, X. (2020). "Simplify the usage of lexicon in Chinese NER," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5951–5960. doi: 10.48550/arXiv.1908.05969

Ratinov, L., and Roth, D. (2009). "Design challenges and misconceptions in named entity recognition," in *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, Boulder, USA. 147–155. doi: 10.3115/1596374.1596399

Yan, H., Deng, B., Li, X., and Qiu, X. (2019). TENER: adapting transformer encoder for named entity recognition. *arXiv* [Preprint]. doi: 10.48550/arXiv.1911.04474

Yang, J., Teng, Z., Zhang, M., and Zhang, Y. (2016). "Combining discrete and neural features for sequence labeling," in *Proceedings of the 17th International Conference*, Konya Turkey, 140–154. doi: 10.48550/arXiv.1708.07279

Ye, Y., Wu, Q., Yan, L., Chow, K. P., Hui, L., and Yiu, S. M. (2013). Unknown Chinese word extraction based on variety of overlapping strings. *Inf. Process. Manag.* 49, 497–512. doi: 10.1016/j.ipm.2012.09.004

Zhang, J., Guo, M., Geng, Y., Li, M., Zhang, Y., and Geng, N. (2021). Chinese Named entity recognition for apple diseases and pests based on character augmentation. *Comput. Electron. Agric.* 190, 106464. doi: 10.1016/j.compag.2021.106464

Zhang, Y., and Yang, J. (2018). "Chinese NER using lattice LSTM," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, 1554–1564. doi: 10.48550/arXiv.1805.02023

Zhao, S., Cai, Z., Chen, H., Wang, Y., and Liu, A. (2019). Adversarial training based lattice LSTM for Chinese clinical named entity recognition. *J. Biomed. Inform.* 99, 103290. doi: 10.1016/j.jbi.2019.103290

Zhao, P., Zhao, C., Wu, H., and Wang, W. (2021). Named entity recognition of Chinese agricultural text based on attention mechanism. *Trans. Chin. Society Agric. Machinery* 52 (1), 185–192. doi: 10.6041/j.issn.1000-1298.2021.01.021

Zheng, J. H., and Wen-Hua, L. I. (2002). A study on automatic identification for Internet new words according to word-building rule. *J. Shanxi Univ. (Natural Sci. Edition)* 25, 115–119. doi: 10.13451/j.cnki.shanxi.univ(nat.sci.).2002.02.007

Genetic diversity assessment of *Hopea hainanensis* in Hainan Island

Yukai Chen[1†], Hai-Li Zhang[2†], Li Zhang[3*],
Mir Muhammad Nizamani[4*], Taoxiu Zhou[5], Haiyang Zhang[6]
and Tingting Liu[3]

[1]Ministry of Education Key Laboratory for Ecology of Tropical Islands, College of Life Sciences, Hainan Normal University, Haikou, China, [2]Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, School of Life Sciences, Hainan University, Haikou, China, [3]Guizhou Normal University Museum, Guizhou Normal University, Guizhou, China, [4]Department of Plant Pathology, Agricultural College, Guizhou University, Guiyang, China, [5]College of Biological Science and Technology, Yangzhou University, Yangzhou, China, [6]College of International Studies, Sichuan University, Chengdu, China

Hopea hainanensis (Dipterocarpaceae) is an endangered tree species restricted to Hainan Island, China, and a small part of Northern Vietnam. On Hainan Island, it is an important indicator species for tropical forests. The wood of *Hopea hainanensis* has a very high utilization value in nature since it is compact in structure, hard in texture, not easily deformed after drying, durable, and resistant to sunlight and water. As a result of its high quality, it has been felled and mined by humans without restraint, resulting in a reduction of its population size, severe habitat fragmentation, and a sharp decline in its population. Therefore, its conservation biology needs to be researched urgently. Researchers are currently focusing on the ecological factors and seed germination in the habitat of *Hopea hainanensis* to determine its endangered status. In the literature, there are no systematic analyses of the endangered mechanism of *Hopea hainanensis* in terms of genetic diversity. It focuses especially on the systematic genetic diversity of *Hopea hainanensis* in fragmented habitats. Using single nucleotide polymorphism (SNP) and genotyping-by-sequencing (GBS) technology, 42 samples from seven different cohabitation groups were genotyped. The results showed that the average heterozygosity of the seven populations of *Hopea hainanensis* was 19.77%, which indicated that the genetic diversity of *Hopea hainanensis* was low. Genetic diversity research is essential for rare and endangered plant protection research. We can find a scientific basis for protecting endangered plants on slope bases by analyzing genetic differences and relationships among populations.

KEYWORDS

conservation biology, endangered tree, hopea hainanensis, SNP, GBS

1 Introduction

The Earth's biodiversity quickly decreases due to agricultural growth, over-exploitation, deforestation, pollution, and climate change (Wang et al., 2007). Around 40% of plant species are on the verge of extinction (Ly et al., 2018). Conservation genetics, a new science that uses population genetics principles and methods to biological conservation, aims to save endangered species from extinction (Hogarth et al., 1997; Sarath Padmanabhan and Hemaprabha, 2018). Endangered animals are frequently distinguished by tiny, scattered populations and limited gene flow among populations (Mehmood et al., 2018). Mating happens more commonly among relatives in tiny, isolated populations, and selfing may be observed in hermaphroditic plants. Inbreeding causes homozygosity in harmful recessive genes and, as a result, the generation of poorer offspring, a condition known as inbreeding depression (Kardos et al., 2016).

Furthermore, genetic drift is higher in small populations, contributing to the fixation of harmful mutations and loss of genetic variation, weakening a population's adaptive ability and raising its extinction risk (Kardos et al., 2021). The area of conservation genetics, which aims to research the genetic diversity, population differentiation, mating system, and historical demography of endangered species, gives amazing insights into preserving biodiversity in the real world (Brown et al., 1991). Furthermore, *Hopea hainanensis* research is primarily concerned with the impacts of various environmental conditions in the habitat on seed germination and seedling development, *ex situ* conservation, and better seed selection and cultivation techniques in artificial propagation and cultivation (Zhang et al., 2022).

Dipterocarpaceae comprise 20-50% of the basal forest area and more than 50% of the canopy trees in tropical Asian forests (Ghazoul, 2016). Because many Dipterocarpaceae species are valuable wood resources, they have been widely exploited in tropical Asian nations. As a result of the widespread harvesting of timber and destruction for agriculture, many dipterocarps are now designated as endangered or severely endangered (Wang et al., 2021). On the other hand, Dipterocarp woods are considerably more than just a supply of lumber. They are vital components of Asian tropical rainforest ecosystems, acting as the foundation for these varied ecosystems. Indeed, Southeast Asia is home to four of the world's 25 "biodiversity hotspots" (Wang et al., 2020). Furthermore, dipterocarp forests provide various ecosystem services and play a significant role in balancing ecological processes at the regional and global levels (Agoramoorthy, 2002). It is the representative and endemic species of the tropical ravine rainforest in Hainan.

The natural survival population of *H. hainanensis* in Hainan is mainly distributed in the forest patches dominated by broken secondary rainforests in and around Limushan, Bawangling, Jianfengling, Diaoluoshan, Yinggeling, Wuzhishan, and other forest areas in Hainan Island (Song et al., 2020). *H. hainanensis* is listed as a class I protected plant in the List of China's National Key Protected Wild Plants (Jiang, 2019). It was identified as an endangered species in the Red Book of Chinese Plants and is ranked as "Endangered" by the IUCN (Ly et al., 2018). Currently, studies on *H. hainanensis* mainly focus on the effects of various environmental factors in the habitat on seed germination and seedling development, *ex situ* protection, and improved seed selection and cultivation techniques in artificial propagation and cultivation (Chen et al., 2015). There is a lack of systematic analysis of the endangerment mechanism in terms of genetic diversity, and there is even less research on the systematic genetic diversity of ports in different fragmented and chemical habitats (Li et al., 2015; Zhang et al., 2022).

However, genetic diversity as an extinction mechanism for *H. hainanensis* has not been systematically studied. A few studies have been conducted on the genetic diversity of *Hopea hainanensis* systems in fragmented and metaplastic environments. Wang et al. isolated and identified 12 polymorphic microsatellite markers on endangered *H. hainanensis* (Wang et al., 2020). The genetic diversity and population structure of 10 *H. hainanensis* populations were analyzed using 12 SSR markers in Hainan Island. The emergence of population bottlenecks may cause genetic diversity loss and population differentiation. Long-term protection strategies for endangered species in Hainan were proposed.

In many fields, genotyping by sequencing (GBS) in simplified genome sequencing technology has become increasingly popular as next-generation high-throughput sequencing technology has developed (Mehmood et al., 2022). These include the construction and analysis of genetic maps, the study of genome-wide association systems and gene diversity and identifying the germplasm of plants and animals. Therefore, in this study, GBS technology was used to systematically identify 42 genome-wide SNPs of *H. hainanensis* resources. Based on the identified SNPs, a phylogenetic tree of these 42 *H. hainanensis* resources was constructed, and genetic diversity was analyzed. It has practical guiding significance for the protection of *H. hainanensis* resources and the improvement of its ecological environment. It is of great significance to the protection of *H. hainanensis* biodiversity.

2 Materials and methods

2.1 Study area

Hainan Island (E108°37′-111°03′, N18°10′-20°10) is located in southern China (Zhang et al., 2021), and it is the largest island city in China (Figure 1). Hainan Island has a mild climate, with
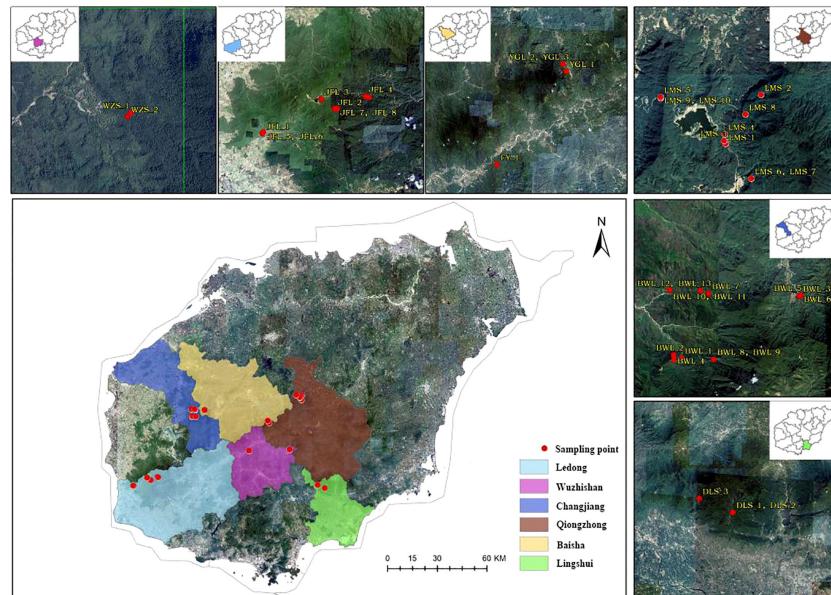
FIGURE 1
A distribution and sample collection information map of *Hopea hainanensis* in Hainan Island.

an annual average temperature of 22-27°C, and is rich in forest resources (Zhang et al., 2022). Hainan Island is low and flat all around and towering in the middle, with Wuzhishanand Yinggeling as the uplift cores and descending step by step to the periphery. (https://www.hainan.gov.cn). Hainan Island is hailed as the largest "natural museum" by biologists, and 102 rare animals in Hainan Island have been included in the list of national first- and second-class key protected animals (Zhang et al., 2022). There are many rare and endangered wild plants. At the same time, northern and coastal regions have relatively low biodiversity and fewer rare and endangered species due to greater human disturbance intensity (Nizamani et al., 2021).

2.2 Materials

In the early field investigation, field investigation and actual sampling possibilities are based on the natural distribution of the levee population. The *H. hainanensis* sample materials were divided into seven regional populations according to their geographical sources for population sampling. The bawangling population includes nine subpopulations, namely BWL P1-P9, including samples BWL1-BWL_13. Limushan population consists of 7 subpopulations, LMS P1 and LMS P3-P8, including samples LMS1-LMS_10, respectively. The Jianfengling population included seven subpopulations JFL X1-X2, JFL P1, JFL P3-P6, which included samples JFL1-JFL_10. Wuzhishan population included two subpopulations

(WZS P1-P2), including samples WZS_1-WZS_2, respectively. Yinggeling population included two subpopulations YGL P1-P2, including YGL1-YGL_3, respectively. The Diaoluoshan population included one subpopulation, DLS P1, including samples DLS_1-DLS_3. Fanyang population included one subpopulation FY1, including sample FY_1. The Fanyang population belongs to the Wuzhishan region. Regional population division, longitude, latitude, and sample number are shown in Table 1.

2.3 Methods

2.3.1 Sample DNA extraction and quality testing

42 *H. hainanensis* leaf samples were extracted with a Biotech (Beijing, China) DNA extraction kit (Plant no. DP305) under the standard operating procedure. After DNA extraction, the quality and concentration of DNA samples should be tested. Qubit was used to determine the concentration of DNA samples, and AGE was used to detect the quality of DNA samples (Wang et al., 2015).

2.3.2 Construction and sequencing of genomic GBS library

The quality and concentration of the extracted DNA were tested to be qualified and then sent to Hangzhou Lianchuan Biotechnology Co., Ltd. for GBS library construction and

TABLE 1 The list of information for *Hopea hainanensis* sample collection in Hannan Island.

Area code	Location	Longitude	Latitude	Sample	Number of sample
BWL	Bawanglin	109°03′–109°17′E	18°57′–19°11′N	BWL_1-BWL_13	13
LMS	Limushan	109°39′–109°48′E	19°07′–19°14′N	LMS_1-LMS_10	10
JFL	Jianfengling	108°44′–109°02′E	18°23′–18°52′N	JFL_1-JFL_10	10
WZS	Wuzhishan	109°39′–109°47′E	18°49′–18°58′N	WZS_1-WZS_2	2
YGL	Yinggeling	109°11′–109°34′E	18°49′–19°08′N	YGL_1-YGL_3	3
DLS	Diaoluoshan	109°11′–109°34′E	18°49′–19°08′N	DLS_1-DLS_3	3
FY	Fanyang	109°27′E	18°53′N	FY_1	1

sequencing. RsaI-HaeIII digestion was used for digestion. The high-throughput sequencing library was prepared through terminal repair, A-tail addition, sequencing adaptor addition, purification, PCR amplification, and so on. The library was purified by electrophoresis and gluing according to the preset scheme. The gluing range of the library was 450-500bp to select the library with the length of the inserted fragment in the target interval for subsequent sequencing. Only libraries that had been screened for fragment length were qualified for sequencing. The sequencing platform was Illumina Nova6000, and the sequencing mode was PE150.

2.3.3 SNP mining in *H. hainanensis* genome

After the sequencing data is taken off the machine, the raw read data obtained by sequencing is quality-controlled, and low-quality sequences and splice sequences are removed to obtain a clean read. After that, all samples' reads are clustered, similar reads are clustered together as a tag, and the consensus sequence is generated. Then, the data were aligned with the consensus sequence by the individual, and the Clean Read data were aligned with the consensus sequence. GATK software and SAMTOOLS software were used for SNP detection, and the quality filtering of the detected mutation sites was carried out. The evolutionary analysis was based on SNP data. Before evolutionary population analysis, SNPs were filtered according to the minor allele frequency (MAF) > 0.05 and data integrity > 80% (i.e., no more than 20% of individual genotypes were missing).

2.3.4 Phylogenetic tree analysis of *H. hainanensis*

The phylogenetic tree is a diagram used to describe the genetic differentiation relationship between species. According to the evolutionary relationship between different sources and different types of organisms, all kinds of organisms are placed on the branching tree diagram. The evolution process and the relationship between these organisms are succinctly described. Based on SNP, 1000 replicates of the PDIST model were obtained as phylogenetic trees of all samples based on the neighbor-joining algorithm of MEGA software (Dieckmann et al., 2016).

2.3.5 Principal component analysis

Principal Components analysis (PCA) was performed based on SNP to obtain the clustering of Principal components of all samples. Through PCA analysis, it can be known which samples have short genetic distances and close relatives. The samples with long genetic distances and distant relatives are helpful for population genetic evolution analysis.

2.3.6 Analysis of population genetic structure

Population genetic structure analysis can provide information on the origin and composition of individual lineage. Based on SNP, the population structure of all samples was analyzed by ADmixture software, and the number of clusters (K value) was assumed to be 1-10, respectively. Different K values represent the distribution of ancestral genetic material of different populations under the assumption that there are K ancestral groups.

2.3.7 Analysis of the genetic relationship

Based on SNP, the genetic distance between pairs of all samples was calculated. We can get the relative distance between samples by analyzing genetic distance data, which can assist the evolution analysis. In the phylogenetic heat map, the redder the color, the closer the relationship between the two individuals on the horizontal and vertical axes, the large area of red in the phylogenetic heat map between multiple individuals indicates that these individuals may constitute a closely related family group. Conversely, the bluer the heat map, the more distant the relatives.

3 Results

3.1 The quality of sequencing

After GBS sequencing, 28.09 Gb of Raw READ data were obtained from 42 *H. hainanensis* samples. After removing low-quality sequences, sequences containing more than 5% N (N represents undetermined base information), and adapter sequences, 27.85 Gb of high-quality sequencing data (Clean data) was obtained. The average size of each sample is 0.66 Gb. The average proportion of base error rate below 1% (Q20)

was 96.66%, and the average proportion of base error rate below 0.1% (Q30) was 91.34%, indicating the high quality of sequencing. The average ratio (GC content) of guanine (G) and cytosine (C) among the four bases of DNA was 47.84%, indicating that the distribution was reasonable. The data overview of each sample is shown in Table 2.

3.2 SNP site mining

After comparing the data to the consensus sequence, GATK and SAMTOOLS software were used for mutation detection (Wright et al., 2019). SNPs consistently output by the two software were retained as reliable loci. According to the

TABLE 2 The materials used in this study and overview of the GBS dataset.

Sample	Raw data (bp)	Raw data	Clean data (bp)	Clean data	Effective data (%)	Q20 (%)	Q30 (%)	GC (%)
BWL_1	3889430	0.58G	3840274	0.57G	98.5	96.77	91.59	50.46
BWL_2	4353376	0.65G	4307204	0.64G	98.77	96.97	92	47.31
BWL_3	4410548	0.66G	4361182	0.65G	98.67	96.65	91.3	48.35
BWL_4	5189486	0.78G	5136018	0.77G	98.78	96.97	91.97	47.43
BWL_5	4267498	0.64G	4218714	0.63G	98.63	96.74	91.53	47.6
BWL_6	4538054	0.68G	4487112	0.67G	98.67	96.75	91.5	47.63
BWL_7	4157322	0.62G	4111528	0.62G	98.68	96.68	91.35	47.67
BWL_8	4475100	0.67G	4426518	0.66G	98.71	96.66	91.3	47.72
BWL_9	4937116	0.74G	4886146	0.73G	98.81	96.68	91.31	47.43
BWL_10	4079886	0.61G	4027066	0.60G	98.5	96.47	90.94	47.86
BWL_11	4272788	0.64G	4224600	0.63G	98.64	96.62	91.23	47.32
BWL_12	3689286	0.55G	3652654	0.55G	98.8	96.89	91.79	47.31
BWL_13	5344202	0.80G	5276592	0.79G	98.52	96.61	91.26	47.54
LMS_1	4132158	0.62G	4078226	0.61G	98.48	96.79	91.65	47.51
LMS_2	4568574	0.69G	4503136	0.67G	98.39	96.3	90.59	49.58
LMS_3	4821746	0.72G	4757702	0.71G	98.48	96.6	91.24	48
LMS_4	4205644	0.63G	4159754	0.62G	98.71	96.84	91.7	49
LMS_5	5239770	0.79G	5175988	0.77G	98.58	96.7	91.44	47.39
LMS_6	4629342	0.69G	4577802	0.69G	98.69	96.89	91.83	47.7
LMS_7	4531974	0.68G	4476688	0.67G	98.59	96.7	91.45	48.19
LMS_8	4704600	0.71G	4654786	0.70G	98.75	96.76	91.49	47.3
LMS_9	4455698	0.67G	4406112	0.66G	98.68	96.77	91.55	47.45
LMS_10	3780250	0.57G	3741060	0.56G	98.75	96.83	91.67	47.82
JFL_1	4217110	0.63G	4166782	0.62G	98.61	96.64	91.27	47.93
JFL_2	5198358	0.78G	5130962	0.77G	98.53	96.46	90.9	47.97
JFL_3	4385206	0.66G	4324702	0.65G	98.44	96.4	90.79	47.67
JFL_4	2491648	0.37G	2453514	0.37G	98.24	96.34	90.71	47.39
JFL_5	4668386	0.70G	4607144	0.69G	98.54	96.26	90.48	47.9
JFL_6	6462998	0.97G	6396564	0.96G	98.8	96.84	91.63	46.6
JFL_7	2447580	0.37G	2423086	0.36G	98.69	96.78	91.57	48.35
JFL_8	5394202	0.81G	5330052	0.80G	98.61	96.71	91.43	47.75
JFL_9	5101784	0.77G	5043916	0.76G	98.67	96.77	91.57	47.92
JFL_10	6088498	0.91G	6023038	0.90G	98.76	96.81	91.6	47.54
WZS_1	4060186	0.61G	4003372	0.60G	98.42	96.69	91.45	47.78
WZS_2	6086906	0.91G	6017660	0.90G	98.67	96.84	91.67	47.33
YGL_1	4555380	0.68G	4505358	0.67G	98.74	96.89	91.82	48.3
YGL_2	4127066	0.62G	4066052	0.61G	98.3	96.18	90.33	47.46
YGL_3	3659212	0.55G	3615938	0.54G	98.63	96.43	90.84	47.64
DLS_1	3968904	0.60G	3921372	0.59G	98.54	96.8	91.64	48.71
DLS_2	3803398	0.57G	3754674	0.56G	98.5	96.49	90.98	48.3
DLS_3	4899390	0.73G	4838058	0.72G	98.58	96.71	91.44	48.09
FY_1	3039210	0.46G	2882090	0.43G	94.63	96.19	90.57	47.28

criteria of MAF >0.05 and data integrity >0.8, SNP data were further processed and filtered to retain SNP$_S$ with polymorphisms. After filtering the SNPs obtained, 430376 high-quality SNPs were finally obtained for subsequent analysis. It can be seen from the following Table 3 that the heterozygosity of the Fanyang population (FY) is relatively high, which may be related to the fact that the Fanyang population has only one sample, the sample size is small, the width of the genetic variation is insufficient and other factors, so there is not enough sample data for comparative analysis of the genetic diversity in this population. The heterozygosity of the other six populations ranged from 19.26% to 20.34%, with average heterozygosity of 19.77%, indicating a low level of genetic diversity.

3.3 Genetic evolution and population analysis

3.3.1 Phylogenetic evolutionary tree

The identified high-quality SNPs were used for phylogenetic analysis of the 42 *H. hainanensis* resources. After 1000 repetitions based on the PDIST model, the neighbor-joining algorithm of MEGA software was used to perform evolutionary analysis on all samples, and the phylogenetic tree of 42 *H. hainanensis* sample resources was obtained (Figure 2). Samples from the same sampling site were relatively closer to each other. However, the relative distance between the samples from different sites means that the internal samples from different sampling sites in these seven population areas may have a common ancestor. The results showed that the 42 samples could be divided into two large groups, and each could be further divided into small subgroups. In general, the samples of the same geographical origin were relatively aggregated in the two large taxa. Still, the distribution was mixed in the small subgroups, and the samples of the same geographical origin were not completely merged into the same subgroup. Group I mainly include the resources from Diaoluoshan, Fanyang Mountain, and Yinggeling, and the resources from Wuzhishanare clustered into Group II. The resources from Limushan, Bawangling, and Jianfengling are distributed in both groups, and the distribution is relatively chaotic. The small subgroups clustered in group I were divided

into three small independent subgroups, indicating a large difference in kinship distance between the large group and each other. The aggregation of samples in group II was relatively uniform. Therefore, although there is certain geographical isolation between the *H. hainanensis* resources of different population areas, there is no direct correlation between the clustering based on genetic distance and its geographical source.

3.3.2 Analysis of population genetic structure

To further verify the evolutionary genetic relationship between the samples and infer that the *H. hainanensis* population likely came from several ancestors. The genetic structure of the mutations in each sequencing sample was further studied. Based on SNP data, ADmixture software was used to analyze the population structure of all samples. Then, cluster analysis was performed, assuming that the number of clusters (K value) was 1-10. Different K values represent the distribution of ancestral genetic material of different populations under the assumption that there are K ancestral groups. Since K=1 cannot represent the distribution of ancestral genetic material of different populations, it is not shown in the figure. As shown in Figure 3, when K=2 and the sample are divided into two subgroups, the sample of group 1 is almost dark blue, and the sample of group 2 is almost light purple. The samples from Fanyang (FY), Diaoluoshan (DLS), Jianfengling (JFL), and Limushan (LMS) were clustered into group 2, and the remaining samples were clustered into group 1. In the Cross-Validation (CV) errors graph (Figure 4), when K=2, CV error achieves the minimum value, indicating that the genetic differences between samples are relatively large and the genetic relationship is distant. Therefore, it can be preliminarily concluded that the seven *H. hainanensis* populations in Hainan Island came from two different ancestors, and there was less gene exchange among them. In the table of genetic differentiation coefficients among populations (Table 4). The F_{st} values among the seven *H. hainanensis* populations ranged from -0.05258 to 0.29542. There was significant genetic differentiation ($F_{st} > 0.25$) between FY, WZS, and DLS populations. The genetic differentiation between DLS and BWL, WZS populations, and FY and BWL populations was significant ($0.15 < F_{st} < 0.25$). There was a moderate genetic differentiation between BWL and

TABLE 3 SNP Statistical results.

Population code	SNP number	Heter LociNum	Homo LociNum	Hetloci-ratio
BWL	84652	16895	67757	19.96%
LMS	99347	19073	80275	19.26%
JFL	86427	17578	68849	20.34%
WZS	92887	18325	74562	19.73%
YGL	80498	15682	64816	19.48%
DLS	86604	17195	69409	19.85%
FY	17510	9285	8225	53.03%
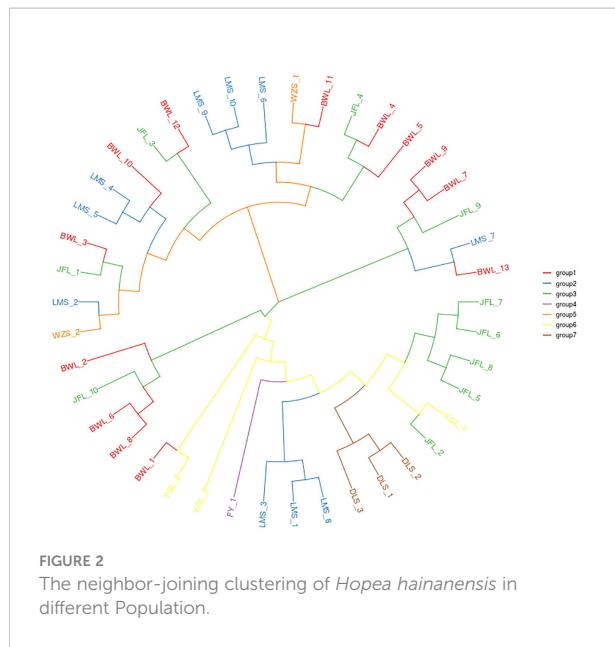
FIGURE 2
The neighbor-joining clustering of *Hopea hainanensis* in different Population.

between groups. Samples with similar genetic backgrounds will gather in the figure to form a cluster. The farther the distance between the two samples in the figure, the greater the genetic background difference between the two samples. As shown in the following figure (Figure 5), the 42 *H. hainanensis* were clustered to form three independent clusters, among which eight samples from Jianfengling (JFL_2, JFL_5-JFL_8), Limushan (LMS_3, LMS_8) and Yinggeling (YGL_1) populations with similar genetic backgrounds were clustered together to form cluster 1. Fanyang (FY_1), Wuzhishan(WZS_1-WZS_2), Bawangling (BWL1-BWL_13), Yingge Mountain (YGL_2, YGL_3), Jianfeng Mountain (JFL_1, JFL_3-JFL_4, JFL_9-JFL_10) and Limushan (LMS_1, LMS_2, LMS_4-L). The 31 samples from the six populations of MS_7, LMS_9, and LMS_10 were clustered together with similar genetic backgrounds to form cluster 2. The population of DLS_1-DLS_3 was far away from the other 2 clusters, showing a long genetic distance, so the population of DLS_1-DLS_3 formed a cluster alone.

3.3.4 Analysis of the genetic relationship

In the relatedness heatmap (Figure 6), the relatedness coefficient was more significant than 0.4 (between the three samples of DLS_1, DLS_2, and DLS_3). The relatedness among the three samples of Diaoluoshan was very close. The genetic distance between LMS_8 and LMS_1 was very close in the relatedness heatmap. It can be concluded that the samples in the same population are more closely related, and the more distant the geographical location, the more complex the gene exchange, and the more distant the genetic relationship. The six samples, YGL_1, JFL_2, and JFL_5-JFL_8, are closely related. The three samples from Diaoluoshan (DLS_1-DLS_3) and

JFL, YGL, FY and LMS, JFL and DLS, and the other three populations (0.05 < Fst < 0.15). In addition, the genetic differentiation among other populations was low, so differentiation could not be considered (F_{st} < 0.05).

3.3.3 Principal component analysis

Principal component analysis (PCA) was carried out on H. hainanensis population samples from Hainan Island to determine the evolutionary relationship among the populations further. When the geographical distance between groups is relatively close, PCA can better reflect the relationship

FIGURE 3
The population structure analysis on *Hopea hainanensis*.

FIGURE 4
K selection of population structure.
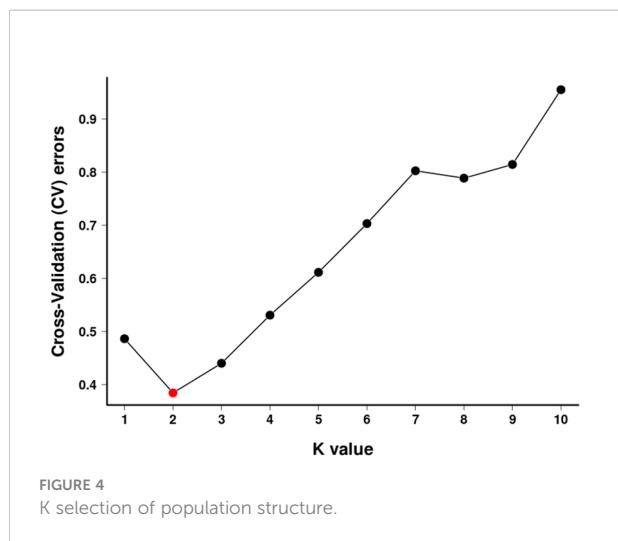
Limushan (LMS_1) are just between 0.2 and 0.3. This indicates that there is still some genetic exchange between clusters under geographical isolation.

4 Discussion

4.1 Genetic diversity in *Hopea hainanensis*

SNP variation is the most important and widespread type of sequence variation in the plant genome, which can be easily identified by sequence alignment (Fang et al., 2014). In this study, 48795 high-quality SNPs were obtained by screening and filtering. In the natural state of *H. hainanensis* field, the ecological range of the population is wide. The seeds are winged nuts, and the germination rate is high. Still, the seeds have higher requirements for germination conditions in the natural environment, which restricts the development of the population (Trang and Triest, 2019). Even if the *H. hainanensis* seeds can germinate and grow into seedlings in the natural population, the *H. hainanensis* seedlings are easy to be eliminated due to their weak competitiveness, resulting in few

remaining adult *H. hainanensis* plants and weak natural regeneration ability of the population in the field (Kenta et al., 2004; Mehmood et al., 2021). The population density was very low, leading to the population's weak reproductive ability and stress resistance and slow natural recovery and development. Genetic diversity is lost when the effective population shrinks and mating is switched from outcrossing to selfing (Ellegren and Galtier, 2016; Cai et al., 2021). It is most likely that a severe demographic bottleneck is responsible for the low genetic diversity of *H. hainanensis* populations. Over the past 300 years, this species has lost about 70% of its population (Ly et al., 2018). In the 20th century, Hainan Island's deforestation increased rapidly. About 80% to 95% of the primary forests have been destroyed because of logging for wood on a large scale. Furthermore, transitions to rubber trees and Eucalyptus plantations, and the growth of cities (Lin et al., 2017; Chen et al., 2018; Sun et al., 2020). Due to the high quality of its wood, the number of *H. hainanensis* trees would go down proportionally, or maybe even more. There is a lack of genetic diversity analysis on the endangered mechanism of levees, especially on the genetic diversity of levees in different fragmented habitats.

Based on SNP, simplified genome sequencing analysis was performed on 42 *H. hainanensis* samples using GBS technology. After obtaining the data, genetic evolution and population analysis were performed, such as phylogenetic tree clustering analysis, population genetic structure analysis, principal component analysis, and phylogenetic relationship analysis. In principal component analysis, the contribution rates of the first principal component (PC1), the second principal component (PC2), and the third principal component (PC3) were 28.78%, 11.2%, and 6.29%, respectively. The contribution rates of the three principal components selected in this analysis were all low, and the total contribution rate was less than 50%. Therefore, there may be a deviation (difference) between the cluster results of PCA and the analysis results of other groups. In the principal component analysis, the genetic distance of the Diaoluoshan population was far from the other populations, and a single cluster was formed. Except for principal component analysis, the population structure of all samples, K value selection of population structure, and phylogenetic evolutionary tree

TABLE 4 Genetic differentiation coefficient(F_{st}: above diagonal).

	BWL	LMS	JFL	FY	WZS	YGL	DLS
BWL		0.02412	0.07731	0.19692	-0.02501	0.06426	0.19716
LMS			0.03763	0.07358	-0.04653	-0.0011	0.11088
JFL				0.06496	0.00063	-0.03522	0.09211
FY					0.25587	0.02041	0.29542
WZS						-0.05258	0.18954
YGL							0.1024
DLS							
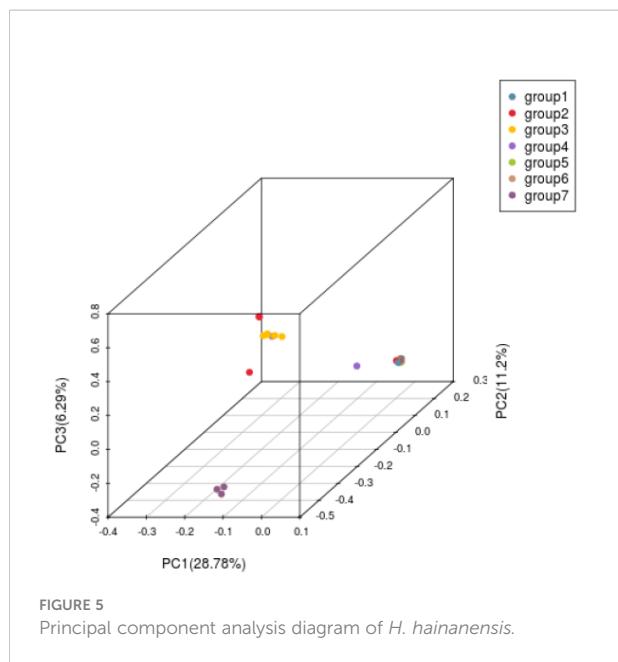
analysis results showed that the cluster division was the same and supported the division of seven populations into two populations. Therefore, it is more reasonable to divide the 42 *H. hainanensis* samples from seven populations into two groups: Group 1 (Diaoluoshan, Limu shan, Yinggeling, Jianfengling) and Group 2 (Wuzhishan, Fanyang, Bawangling). In this study, high-throughput GBS sequencing was performed based on SNPs, and the analysis results may be limited due to the lack of reference genomes covering the whole genome of SNP.
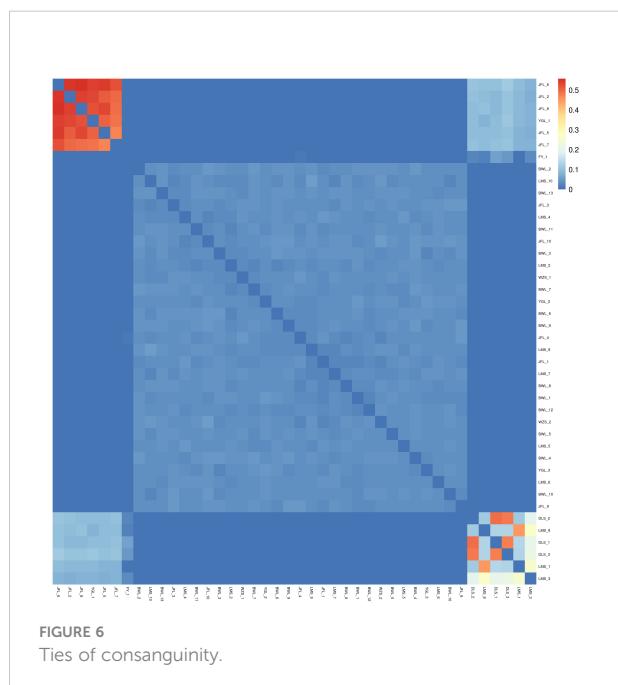
4.2 Population genetic structure and differentiation in *Hopea hainanensis*

Genetic structure is influenced by many factors, such as breeding system, genetic drift, population size, seed dispersal, gene flow, evolutionary history, and natural selection (Konuma et al., 2000; Mehmood et al., 2020). The terrain of Hainan Island is low, flat ground, and high in the middle. The terrain takes Wuzhishan and Yinggeling as the uplifted core and drops progressively to the periphery. The mountain, hill, platform, and plain form a ring-stratified landform with an obvious cascade structure. The samples collected in this study were taken from Wuzhishan, Yinggeling Mountain, and adjacent forest reserves. In geographical location, the Jianfenglin population and Diaoluoshan population, Limushan between groups are far apart (> 100 km). Still, the smaller the genetic distance between the two groups (F_{st} = 0.09211), the existing gene flow between populations may have originated from the common ancestor of genetic exchange, carried by man-made factors, animals or other factors such as geological factors into the other group.

5 Conclusion

In order to improve genetic diversity among *H. hainanensis* populations, the *H. hainanensis* population resources of endangered plants should be effectively protected and developed. In order to protect *H. hainanensis* species, *H. hainanensis* seedlings may be protected *ex situ* due to their weak competitive ability and easy inhibition by mother trees. By conserving *H. hainanensis* seedlings *ex-situ*, we can reduce competition within the population and increase competition between poke stack populations. Genetic drift can also be reduced by increasing gene flow among small populations. Additionally, cross-introduction and breeding among the seven populations can improve genetic diversity.

6 Implications for conservation

Because the loss of genetic variation is a major threat to endangered species, preserving and restoring genetic variation is an important conservation action (Jiang et al., 2018; Cai et al., 2021). We discovered that genetic variation in the populations BWL, WZS and FY were low. These populations are more vulnerable to biotic and abiotic stresses, their conservation is critical. Furthermore, the populations DLS, YGL, JFL and LMS had higher levels of genetic diversity and contained more than one genetic subgroup. That populations could be used as seed sources for propagating seedlings and saplings in restoring Hainan Island's previously logged lowland rainforests. It is

difficult to regenerate native *H. hainanensi*s populations because seedlings and saplings grow slowly and are frequently unable to establish themselves in heavily shaded conditions. To help restore endangered *H. hainanensis* populations on Hainan Island, select populations with high genetic diversity (e.g., for seedlings).

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

Conceptualization, LZ, MMN; methodology, YKC, HLZ, XTZ; software, HLZ, YKC, and MMN; validation, MMN; formal analysis, HLZ, LZ and MMN; investigation, YKC, and XTZ; resources, YKC; data curation, YKC; writing—original draft preparation, LZ, HLZ, HYZ, TTL, XTZ, YKC, and MMN; writing—review and editing, LZ, HLZ, HYZ, TTL, XTZ, YKC, and MMN; visualization, YKC; supervision, YKC; project administration, YKC; funding acquisition, YKC All authors have read and agreed to the published version of the

manuscript. The data presented in the study are deposited in the Dryad repository.

Conflict of interest

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

Publisher's note

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

References

Agoramoorthy, G. (2002). Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions. *J Mammal.*. 83 (2), 630–33.

Brown, A. H. D., Clegg, M. T., Kahler, A. L., and Weir, B. S. (1991). "Plant population genetics," in *Breeding, and genetic resources* (Sinauer, Sunderland), 43–63.

Cai, C., Xiao, J., Ci, X., Conran, J. G., and Li, J. (2021). Genetic diversity of horsfieldia tetratepala (Myristicaceae), an endangered plant species with extremely small populations to China: Implications for its conservation. *Plant Syst Evol.* 307 (4), 1–12. doi: 10.1007/s00606-021-01774-z

Chen, H. X., Huang, C. T., He, F., Zheng, W., and Feng, J. P. (2015). Review on research progress of *Hopea hainanensis. Trop. Forestry* 43 (4), 4–6. doi: 10.1007/s11274-008-9720-8

Chen, B., Xiao, X., Wu, Z., Yun, T., Kou, W., Ye, H., et al. (2018). Identifying establishment year and pre-conversion land cover of rubber plantations on hainan island, China using landsat data during 1987–2015. *Remote Sens.* 10 (8), 1240. doi: 10.3390/rs10081240

Dieckmann, R., Hammerl, J. A., Hahmann, H., Wicke, A., Kleta, S., Dabrowski, P. W., et al. (2016). Rapid characterisation of klebsiella oxytoca isolates from contaminated liquid hand soap using mass spectrometry, FTIR and raman spectroscopy. *Faraday discussions* 187, 353–375. doi: 10.1039/C5FD00165J

Ellegren, H., and Galtier, N. (2016). Determinants of genetic diversity. *Nat. Rev. Genet.* 17 (7), 422–433. doi: 10.1038/nrg.2016.58

Fang, W. P., Meinhardt, L. W., Tan, H. W., Zhou, L., Mischke, S., and Zhang, D. (2014). Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. *Horticulture Res.* 1. doi: 10.1038/hortres.2014.35

Ghazoul, J. (2016). *Dipterocarp biology, ecology, and conservation* (Oxford University Press). https://books.google.com/books

Hogarth, D. M., Cox, M. C., and Bull, J. K. (1997). "Sugarcane improvement: past achievements and future prospects," in *Crop improvement for the 21st century*, *1st edn.* Ed. M. S. Kang (Trivandrum: Research Signpost), 29–56.

Jiang, Z. G. (2019). China's key protected species lists, their criteria and management. *Biodiv Sci.* 27 (6), 698–703. doi: 10.17520/biods.2019089

Jiang, Y., Liu, X., Xiang, W., Jiang, Y., and He, Y. (2018). Genetic diversity and structure analysis of the endangered plant species horsfieldia hainanensis merr. in China. *Biotechnol. Biotechnol. Equip.* 32 (1), 95–101. doi: 10.1080/13102818.2017.1391122

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., et al. (2021). The crucial role of genome-wide genetic variation in conservation. *Proc. Natl. Acad. Sci.* 118 (48), e2104642118. doi: 10.1073/pnas.2104642118

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., and Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Appl.* 9 (10), 1205–1218. doi: 10.1111/eva.12414

Kenta, T., Isagi, Y., Nakagawa, M., Yamashita, M., and Nakashizuka, T. (2004). Variation in pollen dispersal between years with different pollination conditions in a tropical emergent tree. *Mol. Ecol.* 13 (11), 3575. doi: 10.1111/j.1365-294X.2004.02345.x

Konuma, A., Tsumura, Y., Lee, C. T., Lee, S. L., and Okuda, T. (2000). Estimation of gene flow in the tropical-rainforest tree Neobalanocarpus heimii (Dipterocarpaceae), inferred from paternity analysis. *Mol. Ecol.* 9 (11), 1843–1852. doi: 10.1046/j.1365-294x.2000.01081.x

Li, G. Y., Chen, F. F., and Yang, Z. L. (2015). Studies on seedling cultural techniques of Hopea hainanensis. *Trop. Fores.* 43, 7–9. doi: 10.3969/j.issn.1672-0938.2015.04.002

Lin, S., Jiang, Y., He, J., Ma, G., Xu, Y., and Jiang, H. (2017). Changes in the spatial and temporal pattern of natural forest cover on hainan island from the

1950s to the 2010s: Implications for natural forest conservation and management. *PeerJ* 5, e3320. doi: 10.7717/peerj.3320

Ly, V., Nanthavong, K., Pooma, R., Hoang, V. S., Khou, E., and Newman, M. F. (2018). Hopea hainanensis. *IUCN Red List Threatened Species*, 2018–2011. doi: 10.2305/IUCN.UK.2018-1.RLTS.T32357A2816074.en

Mehmood, S., Ahmed, W., Alatalo, J. M., Mahmood, M., Imtiaz, M., Ditta, A., et al. (2022). Herbal plants-and rice straw-derived biochars reduced metal mobilization in fishpond sediments and improved their potential as fertilizers. *Sci. Total Environ.* 826, 154043. doi: 10.1016/j.scitotenv.2022.154043

Mehmood, S., Ahmed, W., Ikram, M., Imtiaz, M., Mahmood, S., Tu, S., et al. (2020). Chitosan modified biochar increases soybean (Glycine max l.) resistance to salt-stress by augmenting root morphology, antioxidant defense mechanisms and the expression of stress-responsive genes. *Plants* 9 (9), 1173. doi: 10.3390/plants9091173

Mehmood, S., Ahmed, W., Rizwan, M., Imtiaz, M., Mohamed Ali Elnahal, A. S., Ditta, A., et al. (2021). Comparative efficacy of raw and HNO3-modified biochar derived from rice straw on vanadium transformation and its uptake by rice (Oryza sativa l.): Insights from photosynthesis, antioxidative response, and gene-expression profile. *Environ. pollut.* 289, 117916. doi: 10.1016/j.envpol.2021.117916

Mehmood, S., Saeed, D. A., Rizwan, M., Khan, M. N., Aziz, O., Bashir, S., et al. (2018). Impact of different amendments on biochemical responses of sesame (Sesamum indicum l.) plants grown in lead-cadmium contaminated soil. *Plant Physiol. Biochem.* 132, 345–355. doi: 10.1016/j.plaphy.2018.09.019

Nizamani, M. M., Harris, A. J., Cheng, X. L., Zhu, Z. X., Jim, C. Y., and Wang, H. F. (2021). Positive relationships among aboveground biomass, tree species diversity and urban greening management in tropical coastal city of haikou. *Ecol. Evol.* 11, 12204–12219. doi: 10.1002/ece3.7985

Piry, S., Luikart, G., and Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *J. Hered* 90 (4), 502–503. doi: 10.1093/jhered/90.4.502

Sarath Padmanabhan, T. S., and Hemaprabha, G. (2018). Genetic diversity and population structure among 133 elite genotypes of sugarcane (Saccharum spp.) for use as parents in sugarcane varietal improvement. *3 Biotech.* 8 (8), 1–12. doi: 10.1007/s13205-018-1364-2

Song, Y. B., Shen-Tu, X. L., and Dong, M. (2020). Intraspecific variation of samara dispersal traits in the endangered tropical tree *Hopea hainanensis* (Dipterocarpaceae). *Front. Plant Sci.* 11, 599764. doi: 10.3389/fpls.2020.599764

Sun, R., Wu, Z., Chen, B., Yang, C., Qi, D., Lan, G., et al. (2020). Effects of land-use change on eco-environmental quality in hainan island, China. *Ecol. Indic.* 109, 105777. doi: 10.1016/j.ecolind.2019.105777

Trang, N. T. P., and Triest, L. (2019). The genetics structure of three threatened Hopea species (Dipterocarpaceae) in the protected areas of Vietnam. *Int. J. Appl. Nat. Sci.* 8 (3), 191–204. doi: doi.org/10.4238/gmr.15028103

Wang, X., Li, J., and Li, Y. (2015). Isolation and characterization of microsatellite markers for an endemic tree in East Asia,Quercus variabilis(Fagaceae). *Appl. Plant Sci.* 3 (6), 1500032. doi: 10.3732/apps.1500032

Wang, C., Ma, X., Ren, M., and Tang, L. (2020). Genetic diversity and population structure in the endangered tree *Hopea hainanensis* (Dipterocarpaceae) on hainan island, China. *PloS One* 15 (11), e0241452. doi: 10.1371/journal.pone.0241452

Wang, C., Ma, X., and Tang, L. (2021). Isolation and characterization of twelve polymorphic microsatellite markers in the endangered *Hopea hainanensis* (Dipterocarpaceae). *Ecol. Evol.* 11 (1), 4–10. doi: 10.1002/ece3.7077

Wang, F. G., Zhang, R. J., Xing, F. W., Ng, S. C., Ye, Y. S., and Chen, H. Q. (2007). Rare and endangered plants and conservation in yinggeling nature reserve, hainan. *J. Wuhan Botanical Res.* 25 (3), 303–309. doi: 10.3969/j.issn.2095-0837.2007.03.017

Wright, B., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., and Grueber, C. E. (2019). From reference genomes to population genomics: Comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics* 20 (1), 1–10. doi: 10.1186/s12864-019-5806-y

Zhang, H. L., Cubino, J. P., Nizamani, M. M., Harris, A. J., Cheng, X. L., Da, L., et al. (2022). *Wealth and land use drive the distribution of urban green space in the tropical coastal city of Haikou, China* (Urban Forestry & Urban Greening), 127554. doi: 10.1016/j.ufug.2022.127554

Zhang, C., Sun, Z., Xing, Q., Sun, J., Xia, T., and Yu, H. (2021). Localizing indicators of SDG11 for an integrated assessment of urban sustainability–a case study of hainan province. *Sustainability* 13 (19), 11092. doi: 10.3390/su131911092

Zhang, L., Zhang, H. L., Nizamani, M. M., Zhou, Q., Su, X., and Chen, Y. (2022). Analyses of community stability and inter-specific associations between a plant species with extremely small populations (Hopea hainanensis) and its associated species. *Front. Ecol. Evol.* 872. doi: 10.3389/fevo.2022.922829

Research on cassava disease classification using the multi-scale fusion model based on EfficientNet and attention mechanism

Mingxin Liu[1], Haofeng Liang[1] and Mingxin Hou[2]*

[1]School of Electronic and Information, Guangdong Ocean University, Zhanjiang, China, [2]School of Mechanical and Power Engineering, Guangdong Ocean University, Zhanjiang, China

Cassava disease is one of the leading causes to the serious decline of cassava yield. Because it is difficult to identify the characteristics of cassava disease, if not professional cassava growers, it will be prone to misjudgment. In order to strengthen the judgment of cassava diseases, the identification characteristics of cassava diseases such as different color of cassava leaf disease spots, abnormal leaf shape and disease spot area were studied. In this paper, deep convolutional neural network was used to classify cassava leaf diseases, and image classification technology was used to recognize and classify cassava leaf diseases. A lightweight module Multi-scale fusion model (MSFM) based on attention mechanism was proposed to extract disease features of cassava leaves to enhance the classification of disease features. The resulting feature map contained key disease identification information. The study used 22,000 cassava disease leaf images as a data set, including four different cassava leaf disease categories and healthy cassava leaves. The experimental results show that the cassava leaf disease classification model based on multi-scale fusion Convolutional Neural Network (CNN) improves EfficientNet compared with the original model, with the average recognition rate increased by nearly 4% and the average recognition rate up to 88.1%. It provides theoretical support and practical tools for the recognition and early diagnosis of plant disease leaves.

KEYWORDS

deep learning, classification, EfficientNet, multi-scale feature fusion, attention mechanism

1 Introduction

With the current climate posing a threat to human health, vegetation and biodiversity (Bhatti et al., 2022a), and the outbreak of the novel coronavirus pneumonia, major cities across the country have suspended production in order to effectively prevent the spread of the epidemic (Bhatti et al., 2022b). The importance of food is self-evident. In recent years, the planting area has continued to expand, and it is also a key food security crop for smallholder farmers because it can withstand harsh conditions. However, with the increase in cassava planting areas, the disease problem is becoming increasingly prominent. Cassava disease can generally be diagnosed according to the shape, color, and leaf shape characteristics of the disease spots on cassava leaves. Under the influence of environmental factors, cassava disease is more likely to occur, which affects the yield and quality of cassava. According to the different characteristics, cassava disease can be mainly classified into Cassava Bacterial Blight (CBB), Cassava Brown Streak Disease (CBSD), Cassava Green Mottle (CGM), Cassava Mosaic Disease (CMD), etc., which can lead to reduced and/or diseased cassava output. Due to the small number of professional plant personnel of cassava and the lack of professional knowledge of general plant personnel of cassava, the symptoms of cassava leaf disease are not typically studied with good understanding, which can lead to inaccurate and incorrect diagnoses of cassava disease. At the same time, the artificial diagnosis and treatment of cassava leaf diseases not only wastes a lot of manpower and material resources but also results in omission and error from subjective judgment due to the relatively similar characteristics of each leaf disease. The diversity of solutions for cassava disease often prevents effective treatment of cassava, so it is very important to correctly identify the disease. Therefore, as an auxiliary means, computer technology can be applied to help planting personnel determine whether there is cassava disease and of which type, and then the best treatment can be given, to avoid yield decline.

There have been many research achievements in judging plant diseases through traditional machine learning methods. Since they are all based on artificial designs of features, they are inefficient and have a large workload (Pujari et al., 2016). In addition, people tend to rely on experience when selecting features, which is highly subjective and not only consumes manpower but also has a large amount of uncertainty. Combining machine learning with deep learning can solve this problem well. Jamil (Jamil et al., 2021) used artificial neural network (ANN) and support vector machine (SVM) to solve the problem of land classification. When the accuracy of ANN was 82.60% and the accuracy of SVM was 73.66%, they combined the two models and weighted them, and finally the average accuracy reached 86.18%. A CNN can automatically extract image features, greatly reducing the workload, while providing a good research direction for plant disease classification.

Meanwhile, K-nearest Neighbor (KNN) classifier (Bazai et al., 2017) and other algorithms are also studied in deep learning in data analysis scenarios, Bhatti (Bhatti et al., 2021a) proposed a method for edge detection of color images by using Clifford algebra and its subalgebra quaternion in image processing, which improved object detection and classification as well as extraction of other features. Bhatti (Bhatti et al., 2021b) also proposed a spatial spectrum HIS classification algorithm – local similarity Projection Gabor Filtering (LSPGF), which uses the reduced-dimension Convolutional Neural Network based on local similarity projection (LSP) and two-dimensional Gabor filtering algorithm. The performance of the proposed method is compared with other algorithms in the public Host Integration Server (HIS) database, and the overall accuracy is better than all datasets. Based on the data information, Bhatti (Aamir et al., 2021; Bhatti et al., 2022c) uses regression analysis algorithm and path analysis algorithm to extract the relationship between variables and get the relationship between algorithms.

With the proposals of AlexNet (Krizhevsky et al., 2012) and Visual Geometry Group Network (VGG) (Simonyan and Zisserman, 2014), the number of network parameters is greatly reduced and the network is more suitable for complex samples under multiple training times, paving the way for deep learning to be applied in future computer vision. In the same year that VGG was proposed, Google proposed GoogleNet (Szegedy et al., 2015). This network adopts the Inception modular structure, by which convolutional kernels of different sizes are used to capture features of feature maps and expand their receptive fields, and then splice the results into channels. Finally, the accuracy of the network is improved by fusing multiple features. In the following years, the proposal of Resnet (He et al., 2016) residual network provided new ideas for the CNN. With the advent of EfficientNet (Tan and Le, 2019), the model has become more capable at capturing features, and its application in computer vision is developing day by day, especially in plant disease recognition.

Hewitt (Hewitt and Mahmoud, 2018) only used a simple shape feature set to judge and recognize relevant plant leaves, in which the feature set included shape features of original leaves and signal features extracted from different convolution models for recognition and obtained good recognition results. In the face of wheat disease leaf identification and differentiation, (Van Hieu and Hien, 2020) used a variety of classification algorithms to compare the prediction accuracy of various neural networks, among which GoogleNet proved to have the highest accuracy of 98%, more suitable for wheat disease detection. As for the detection and recognition of tea diseases, Lee (2018) used Faster Region Convolutional Neural Network (FR-CNN) and candidate objects proposed by Region Proposal Network (RPN) to detect, identify, and distinguish three kinds of tea diseases, with recognition accuracies up to 63%, 81%, and 64%, respectively. Two major crop damage modes in maize production were evaluated, and three commonly used object

detectors were evaluated. It was concluded that YOLOv2 had better performance and was more suitable for the assessment of maize growth damage (Turkoglu and Hanbay, 2019). The adversarial robustness of the model (Zoran et al., 2020) was significantly improved by adding an attention mechanism, and the robustness was effectively improved by changing the model expansion steps. On the lightweight model, Wang (Wang et al., 2022) proposed an Individualized activity space modeler (IASM) mechanism to improve the accuracy and efficiency of the model, and achieved the classification accuracy of 92.57% on the self-made data set by using Ghostnet and Weighted Boxes Fusion (WBF) structures. In the classification of banana diseases, Narayanan (Narayanan et al., 2022) combined the mixed algorithm of CNN and Fuzzy Support Vector Machine (FSVM) to classify banana diseases, CNN to detect, and FSVM algorithm to strengthen the classification, and finally achieved good results. In terms of attention mechanism, Zhu (Zhu et al., 2021) added an attentional mechanism module combining Convolutional Block Attention Module (CBAM) and ECA-Net module to the model, which improved the accuracy of the model by 3.4%. Zakzouk, S. (Zakzouk et al., 2021) used AlexNet to classify new rice diseases with an accuracy of 99.71%. The accuracy of the results indicated the feasibility of the automatic rice disease classification system. Tang (Tang et al., 2020) proposed a new two-stage Convolutional Neural Network image classification network. InnerMove, a new image data enhancement method, was used to enhance images and increase the number of training samples, so as to improve the generalization ability of the deep CNN model for image classification tasks. There are also many innovative neural network methods on algorithm models that can provide ideas for classification. At present, the 3D Convolutional Neural Network method (Hameed et al., 2022a) is innovatively used to extract the feature information for the data set. This method can solve the problem with the data better than the pixel-based support vector machine classifier. When it comes to the impact of food security on local and global economies, Mazhar (Hameed et al., 2022b) applied the sequential model in deep learning to classify the outer layer air particles through the analysis and characteristics of objects and fusion. Compared with the existing deep learning method of surface landscape, the accuracy rate reaches 98%.

In view of the above mentioned contents and problems, in order to improve the efficiency of cassava disease classification and recognition, this paper uses the deep convolutional neural network method according to the characteristics of crop disease images in real scenes, takes cassava leaf disease images as the research object, and designs a cassava disease classification model based on multi-scale fusion CNN. A multi-scale fusion module is proposed to extract multi-scale information features of images. Focal loss was adopted to reduce the emphasis on most categories caused by data imbalance, and to solve the problem of low classification accuracy for categories with few

samples. CBAM (Woo et al., 2018) module was introduced to obtain key information such as texture and color of cassava leaves, and the result of precise positioning of specific features was achieved.

We enhance and amplify the existing data images and add these images to the existing data set to form new mixed data for training the model. The effectiveness of the proposed model was verified by designing several comparison experiments and comparing them with network models such as Resnet and VGG.

The rest of this article is organized as follows. The Materials and Methods section introduces our proposed cassava disease classification model based on multi-scale fusion Convolutional Neural Network. See the results section for experimental results. Finally, in the conclusion part of the summary of this article.

To summarize, the main contributions of this study are as follows:

(i) a model is developed to recognize cassava disease based on Convolutional Neural Network deep learning.

(ii) the accuracy of cassava disease classification model is evaluated using images taken from nature and artificial enhanced images.

(iii) a lightweight module based on attention mechanism to enhance the classification accuracy of the cassava disease model.

2 Method and materials

2.1 Cassava disease classification model based on multi-scale fusion Convolutional Neural Network

In this paper, a classification model of cassava disease was proposed based on multi-scale fusion (shown in Figure 1). The proposed model adopts Efficientnet-B6 as the backbone network, by which a multi-scale fusion module is designed to improve shallow feature extraction. Through CBAM, the channel and spatial weights of subsequent modules are re-calibrated and the classification capacity of cassava leaf disease is enhanced.

2.1.1 EfficientNet

We generally expand the network scale by increasing network depth D, receptive field W, and resolution R. Compared with AlexNet, VGG network convolution kernels were all replaced with smaller 3×3 convolution kernels (including a few 1×1 convolution kernels), which achieved better training results through deeper network structure. However, with the gradual deepening of network depth, network training becomes more difficult due to the emergence of Vanishing Gradient, over-
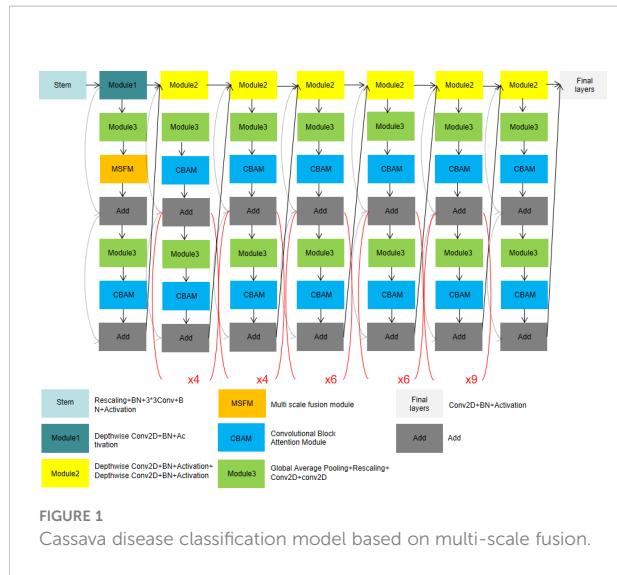
Cassava disease classification model based on multi-scale fusion.

EfficientNet-B6 model structure.

fitting, and other problems. Even if the problem of Vanishing Gradient is solved, the low precision return will lead to high calculation cost and low efficiency of increasing network depth blindly. For another example, MobileNet (Howard et al., 2017) can adjust the number of feature data channels by reducing the amount of computation. However, like deepening the network structure, low precision return will also be found when the width of the network structure reaches a certain level. In the final approach, the neural network can capture finer patterns by using higher-resolution input images. It has developed from 224 × 224 pixels to 229× 229 pixels, or even 512 × 512 pixels. However, accuracy problems are inevitable as the parameter becomes larger. Before EfficientNet, network improvements were generally realized by changing only one of the following variables of network depth, receptive field, and resolution size. However, EfficientNet can obtain better training results by increasing the depth, receptive field, and image resolution through one adjustment. Compared with the aforementioned model, EfficientNet can get a better result because it is capable of adjusting the proportions in three dimensions (shown in Figure 2).

2.1.2 Proposed model
2.1.2.1 Multi-scale fusion module

The low-level feature map has a small stride, a large size, and a small receptive field to detect the details of small targets. The high-level feature map has a larger stride, smaller size, larger receptive field, and rich semantic information. The model extracts detailed features such as color and texture from the low-level network and extracts the blade shape feature from the high-level network.

The key features in the map can be selectively enhanced and the features can be accurately located by redistributing both channel and spatial weights through the Attention Mechanism.
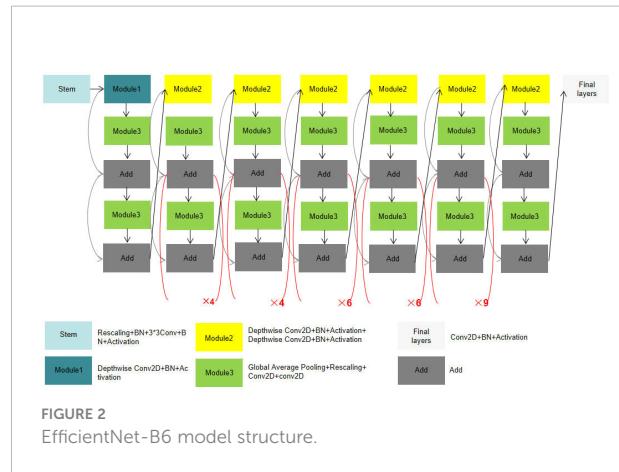
Compared with the previous EfficientNet model, which did not include the MSFM module, the new model adds modules based on the attention mechanism to allocate computing resources to more important tasks. The operations of different pooling layers in channel and space were added to enhance important features and reduce the proportion of unnecessary features. At the same time, the sensitivity field of the low-level feature map augmented by expansion convolution with different expansion rates can not only extract the details of color and texture, but also fully obtain the context information of the image. In the end, the feature map output obtained by different expansion rates is fused to improve the classification effect.

In the result part, we can see from the comparison experiment that our model has higher accuracy compared with other models, and in the heat map, we can see that it has more accurate judgment on key points.

2.1.2.1.1 Convolutional Block Attention Module

The attention mechanism is a relatively efficient data processing method developed in machine learning in recent years, and is widely used in various types of machine learning tasks such as image recognition and natural language processing. When people observe things outside, they usually focus on what they think is important. The attention mechanism focuses on local information that allows the network to achieve better results. Therefore, in this paper, Convolutional Block Attention Module is added before each regularization of the pre-training network and improves the features of the selected maps to increase the accuracy of the model (shown in Figure 3).

2.1.2.1.1.1 Channel Attention Mechanism (CAM)

SENet (Hu J. et al., 2018), as the champion network of the 2017 ImageNet classification Contest, is essentially a model based on a channel attention mechanism, which gives rewards and punishments of different weights according to the importance of each feature channel. In this paper, the channel attention mechanism adopts avg-pool and max-pool for fusion.

FIGURE 3
CBAM structure diagram.

FIGURE 5
Spatial attention mechanism.
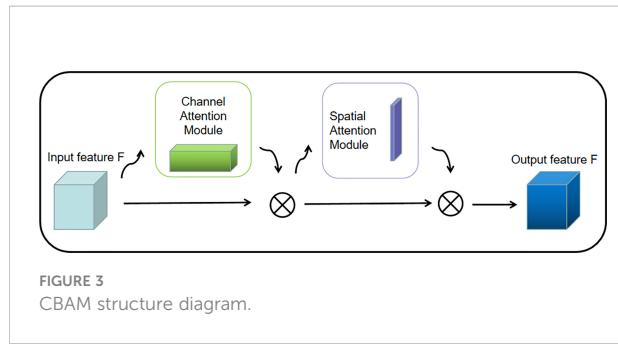
After convolution and activation of the Relu function, the results of the two pooling layers are added together. Finally, it is outputted by the Sigmoid function (shown in Figure 4). The size of the input feature map is H×W×C. Firstly, the global maximum pooling layer and average pooling layer are carried out, respectively, to obtain two-channel weight matrices of 1×1×C.

The two matrix results are fed into a two-layer multilayer perceptron (MLP), respectively, and the MLP can share parameters.

After adding the two feature vectors, the weight coefficients are obtained by the Sigmoid activation function again.

The weight coefficient is multiplied by the original input feature to obtain the final output feature.

2.1.2.1.1.2 **Spatial Attention Mechanism (SAM)**

Different from the weight of each feature plane of the channel attention allocation, the spatial attention model is to find the most important part of the network for processing (shown in Figure 5).

The input characteristic graph is H×W×C. The Max-Pool and Avg-Pool of one dimension are spliced and dimensionally reduced to generate two one-dimensional feature maps.

The weight parameters are generated by the Sigmoid activation function, and then the final output feature is obtained by multiplying the original input feature.

2.1.2.1.2 Multi-scale fusion module

First, the MSFM redistributes the channel and spatial information through the Convolutional Block Attention Module to enhance the characteristics of small target information. Then Feature Extraction is carried out using
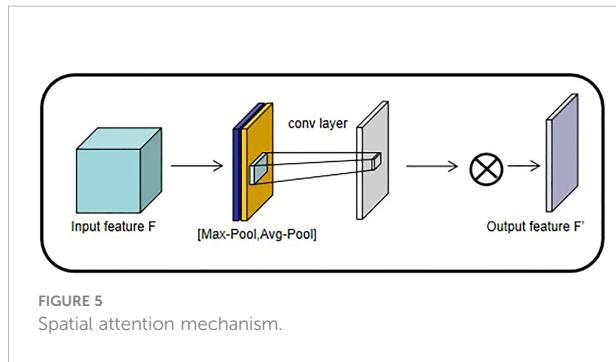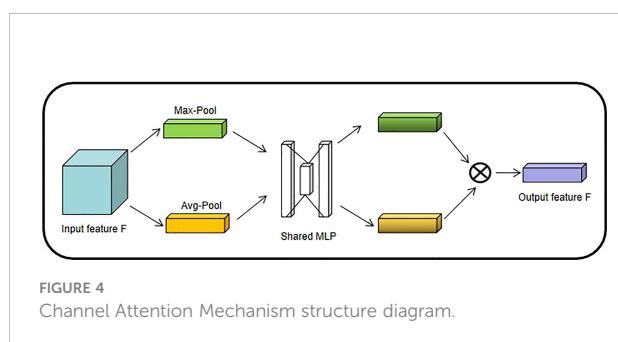
dilated convolution of different dilatation rates, and the context information of feature images is fully extracted by expanding different receptive fields. The convolution kernel size of dilated convolution (Yu and Koltun, 2015) is the same as that of ordinary convolution, and the number of parameters in the neural network remains unchanged. The difference lies in that the dilated convolution has a larger receptive field. A 3x3 convolution kernel with an expansion rate of two has the same receptive field as a 5x5 convolution kernel. However, the number of parameters is only nine, much less than the 25 parameters of the 5x5 convolution kernel. The size of the convolution kernel after expansion:

$$k_d = k + (k - 1) \times (r - 1) \tag{1}$$

where

k_d = Size of the expanded convolution kernel
k = Size of original convolution
r = Dilation rate

The calculation of the receptive field of dilated convolution is as follows:

$$r_f = [(k+1) \times (r-1)+k] \times [(k+1) \times (r-1)+k] \tag{2}$$

where

r_f = Receptive field
k = Size of original convolution
r = Dilation rate

The multi-scale fusion module designed in this paper is shown in Figure 6. First, feature maps are learned by the Convolutional Block Attention Module, and weight calibration

FIGURE 4
Channel Attention Mechanism structure diagram.

FIGURE 6
Multi-scale fusion module.
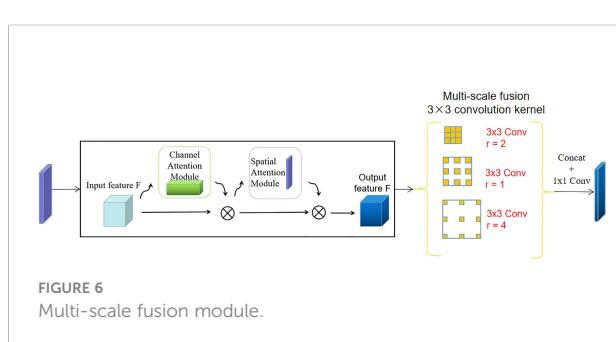
is carried out for channel and space to strengthen the weight reward of important features of feature maps. The invalid features are punished and the weight is reduced to highlight the important information of the detected image. In this paper, a 3×3 convolution kernel is used for feature extraction, and the expansion convolution with expansion rates of 1, 2, and 4 are used for feature extraction, respectively. The features of different scales are extracted from each layer, and the new features are obtained after dimensionality reduction by fusion.

2.2 Loss function

2.2.1 Cross-entropy loss function

In recent years, transfer learning has been widely used in machine learning, which presents satisfactory application results in deep learning. For multi-classification tasks, the cross-entropy loss function (Hu K. et al., 2018) is generally used.

The most commonly used cross-entropy loss function is:

$$CE(p_t) = -a_t log(p_t) \tag{3}$$

where

$CE = Loss$

$P_t = Predictive\ value$

$a_t = Added\ parameters\ that\ represent\ weights\ for\ different\ categories$

Cross entropy loss function under multiple classifications:

$$CE = -\sum_{j=1}^{N} a_t log(p_t) \tag{4}$$

(b) Focal loss

To a certain extent, traditional methods can solve the problem of fewer categories and unbalanced image distribution, but when there are many easily classified samples, the samples will still dominate the training process, so some difficult-to-classify samples have little chance of gaining the attention of the model. The focus function (Lin et al., 2017) treats the difficult-to-classify samples and the easy-to-classify samples differently, focusing on the difficult-to-classify samples and reducing the weight of the easy-to-classify samples. Therefore, the focus function is adopted as the loss function in this study.

Focal loss adjusts the weights of the difficult-to-classify and easy-to-classify samples in the formula:

$$FL(p_t) = -a_t(1 - p_t)^\lambda log(p_t) \tag{5}$$

γ is a constant, and the magnitude of γ determines the weight of small and difficult samples.

When γ<=0, the focusing parameter can be adjusted. When the value of γ is larger, the loss of the sample that is easy to classify is small, and the focus of the model is on the sample that is difficult to classify. This is because when γ is larger, the loss of

small samples and difficult samples will be larger, so that they can obtain greater weight.

When γ>1, the training loss of large and simple samples can be reduced, while the loss of small and difficult samples will not be reduced much.

When γ=1, the equation degenerates into the cross-entropy loss function mentioned above.

The focus loss function of multi-classification:

$$FL = -\sum_{j=1}^{N} (1 - p_t)^\lambda log(1 - p_t) \tag{6}$$

2.3 Input dataset

2.3.1 Data sources and features

The data used in this paper are a cassava leaf dataset manually photographed in Uganda and annotated by experts from the National Crop Resources Institute in collaboration with the Artificial Intelligence (AI) Laboratory at Makerere University, Kampala. The data set contains five kinds of cassava leaf images, and disease images are cassava white leaf blight, brown streak disease, green mottling disease, Mosaic disease and healthy cassava leaf images. It can better reflect the characteristics and symptoms of healthy cassava leaves and diseased cassava leaves in natural environment, and it also represents the real and low diagnostic format that farmers need in real life. The dataset includes images taken under field conditions (some of which are shown in Figure 7). And images enhanced by data. There are 22,000 of them. Each image has a pixel size of 800×600. The standard input size of neural networks such as ResNet、EfficientNet and Resnet-50 is 224×224 pixels. The whole data set was randomly divided into a training set (90%) and a test set (10%). Therefore, 19,800 images were used for model training, and the remaining 2,200 images were used to test the performance of the model.

A Cassava Bacterial Blight (CBB)　**B** Cassava Brown Streak Disease (CBSD)　**C** Cassava Green Mottle (CGM)

D Cassava Mosaic Disease (CMD)　**E** Healthy Cassava Leaf

FIGURE 7
Images of cassava leaf diseases (**A–D** is the disease of four different cassava leaves and e is the healthy leaf).
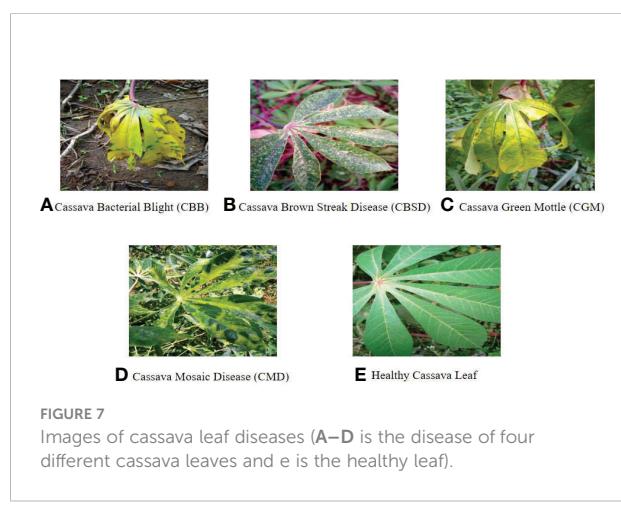
2.3.2 Data augmentation

To prevent network overfitting, OpenCV data enhancement was used to expand the data set appropriately. The data set was enhanced by increasing the brightness, decreasing the brightness, and reversing the image, and the data of various cassava leaves amounted to 4000 pieces. The data enhancement methods used are as follows (shown in Figures 8–12):

Rotation: Rotate the image by 180°.

Brightness reduction: The enhancement factor is 0.7, which means the brightness becomes 70% of the original image.

Horizontal flip: Flip the input image horizontally.

2.4 Computer hardware

The proposed method was tested for training and test configuration for neural network models (shown in Table 1). This result is achieved in Table 2, where PyTorch, as a popular learning framework today, is capable of powerful GPU acceleration and includes deep neural networks. Meanwhile, GPU RTX2080Ti has 11 GB of memory, which can better train the model by adjusting batch size.

2.5 Experimental hyperparameter setting

The cassava leaf data set was divided into a training set, a validation set, and a test set with the ratio of 8:1:1. The training set was trained as an epoch 150 times, 1e-4 was selected as the

FIGURE 8
Cassava Bacterial Blight. **(A)** original **(B)** Rotation by 180° **(C)** Brightness reduction **(D)** Horizontal flipping.

FIGURE 9
Cassava Brown Streak Disease. **(A)** original **(B)** Rotation by 180° **(C)** Brightness reduction **(D)** Horizontal flipping.

FIGURE 10
Cassava Green Mottle. **(A)** original **(B)** Rotation by 180° **(C)** Brightness reduction **(D)**Horizontal flipping.

FIGURE 11
Cassava Mosaic Disease. **(A)** original **(B)** Rotation by 180° **(C)** Brightness reduction **(D)** Horizontal flipping.

FIGURE 12
Healthy cassava leaf. **(A)** original **(B)** Rotation by 180° **(C)** Brightness reduction **(D)** Horizontal flipping.

TABLE 1 Experimental configuration.

Experimental environment	Model and version
Deep learning framework	Pytorch
Programming language	Python3.7
GPU	NVIDIA GeForce RTX 2080 Ti
The hardware environment	Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz 2.10 GHz

initial value of the learning rate in the form of a small amount, and the batch size was set to 16. Batch size not only affects the efficiency of the training model but also affects the accuracy. To find a group balance between efficiency and memory capacity, the batch size is used to calculate the batch size.

2.6 Model evaluation criteria

Accuracy (%) is used as an evaluation index for multi-classification problems in the laboratory. The accuracy of the

TABLE 2 Classification accuracy of the test set in different models.

Model	Accuracy of the test set/%
VGG-16	70.5
Resnet-18	83.2
Resnet-34	83.9
Resnet-50	85.2
Inception v2	84.1
Inception v3	85.3
MobileNet v3	84.1
ShuffleNet	83.5
EfficientNet-b3	85.5
EfficientNet-b6+	86.5
Our model	88.1

experimental model classification can be obtained by removing the number of labels in the test set according to the evaluation criteria. The calculation formula of indicators is as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\,\% \qquad (7)$$

TP=Positive sample prediction is the number of positive classes.
FP=Negative sample prediction is the number of negative classes.

3 Result

3.1 Validation and comparison of proposed Convolutional Neural Network

In order to verify that the improved model in this paper has better image recognition ability compared with the traditional model, this paper uses multiple groups of comparative experiments. The experimental results are shown in Table 2. According to the evaluation index results, as shown in the figure above, the improved model is compared with the model with better performance recently. Among them, the accuracy rate of the model proposed in this paper is 88.1% in the test set separated from the data set, both of which are better than the previous models. To verify that the improved model in this paper has better image recognition ability compared with the traditional model, multi-group comparative experiments were carried out (shown in Table 2).

3.2 Ablation experiments

The confusion matrix (Song et al., 2015) is also one of the evaluation indicators of the classification model (shown in Figure 13). The confusion matrix parameters are converted by classification report. The parameters are given as follows:

(i)The percentage of the total that all predictions are correct, as in Equation 1:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (8)$$

(ii)The proportion of those correctly predicted to be positive that are actually positive, as in Equation 2:

$$Recall = \frac{TP}{TP + FN} \qquad (9)$$

(i)The proportion of correct positive precisions to total positive precisions, as in Equation 3:

$$Precision = \frac{TP}{TP + FP} \qquad (10)$$

The column labels of the confusion matrix represent the predicted cassava leaf disease type, and the sum of the corresponding row values represents the sum of the samples of this type. The diagonal line indicates the number of correctly predicted labels. Each value on the diagonal line indicates the number of correctly predicted labels. The value at the intersection of the columns represents the value of the corresponding tag predicted. If it is not on the diagonal, it can be seen as the number of wrongly predicted tags. The darker diagonal suggests the better model. In the classification results, the judgment accuracy is high, and most of the results of the test set are concentrated on the diagonal of the confusion matrix. The identification accuracy of all kinds of blades is greater than 90%. However, the identification accuracy of Cassava Mosaic Disease in the test set is lower than that of other diseases, and it is easy to misjudge it as other cassava leaf diseases. By observing the confusion matrix, it can be found that the pictures of Cassava Mosaic Disease can be easily identified as Cassava Bacterial Bligh and Cassava Green Mottle, because the symptoms of these three diseases are relatively similar, so classification errors are prone to occur. The obfuscation matrix of the improved model presented in this paper has better performance and a higher average recognition rate.

3.3 Visual output comparative analysis

Class Activation Mapping (CAM) was used to visualize each trained model (Zhou et al., 2016) to better compare the expression process of image features between an improved network and a traditional network. Feng (Feng et al., 2022) used the Grad-CAM thermal map of interpretative analysis, the feature extraction effect of the model can be better expressed. In visualization, the thermal map and the original image are superimposed (shown in Figure 14). This is a visual output of the original image and Efficientnet-B6 and our model, respectively. The darker the color is, the larger the value and the more feasible it will be to serve as the judgment. Compared with the model in this paper, our model has a stronger feature extraction ability and better effects in the face of Cassava leaf disease, which not only extracts different colors of cassava leaf disease but also better captures features in the context.

FIGURE 13
Confusion matrix for the test set.

FIGURE 14
Visual comparison of EfficientNet-B6 and our model.
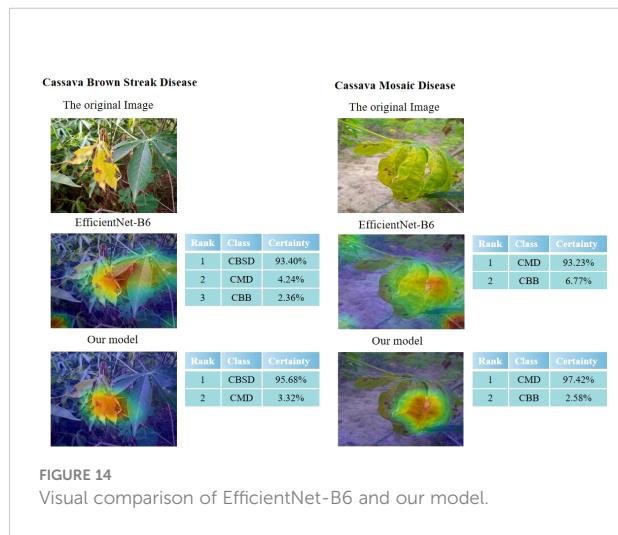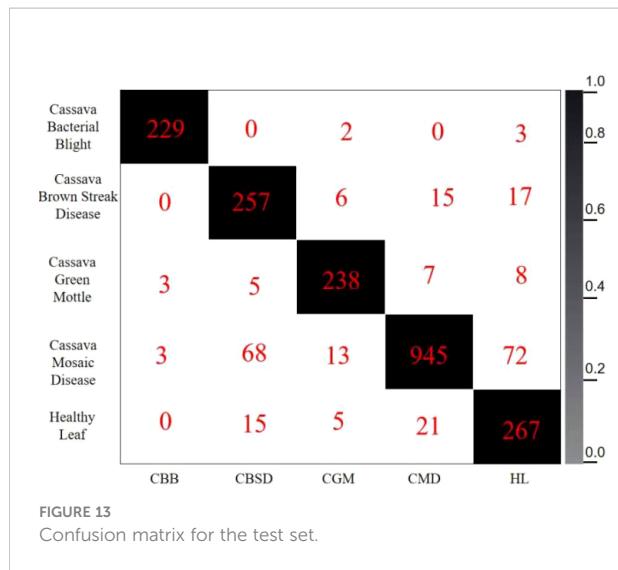
In addition, it can achieve more accurate extraction of key information. It can also be seen from the comparative experiment that the original model has some judgment errors in the color discrimination of leaves and the discrimination ability of background information, while the improved one has a better grasp of sample information and a better capture effect for judgment features.

4 Conclusions

Cassava, as one of the three varieties of potato, is an important food. Compared with the diagnosis of cassava disease by human, the identification of cassava disease by computer and deep learning method not only has low cost and higher accuracy than manual diagnosis, but also greatly reduces the efficiency. The multi-scale cassava leaf classification model proposed in this paper can better ensure the safety and efficiency of cassava food production and judge the cassava disease type more accurately. Compared with the diagnosis of cassava leaf disease using the human eye, the identification of cassava leaf disease by computer and deep learning method is characterized by lower cost, higher accuracy, and greatly increased efficiency. In this paper, a multi-scale fusion module was proposed, and the Focal Loss function and CBAM module were introduced. An optimization network model of a multi-scale fusion network based on EfficientNet and attention mechanism was proposed. The model was used to train the cassava leaf disease data set and compare with the EfficientNet, ResNet50, and VGG16 networks. The experimental results show that the improved network proposed in this paper has higher precision and better generalization ability. The problem of uneven data was solved by changing the loss function, the distinguishing ability of cassava leaf diseases was improved through an attention mechanism, and the recognition ability of the model was enhanced by multi-layer fusion. According to the pricing standard of the model, the model proposed in this paper can be used for image recognition of Cassava leaf disease.

Since our model adds the lightweight module MSFM based on EfficientNet, our model can be installed on mobile devices, such as microprocessors. Due to the large-scale application of 5th Generation Mobile Communication Technology (5G) (Johannes et al., 2017), there is efficient transmission. According to the improvement of the hardware configuration of mobile terminal equipment, the image to be detected can be uploaded to the cloud server for processing, and then the recognition and classification results can be returned to the terminal. For some cassava planting technicians, when they have doubts about cassava disease judgment, the mobile terminal deployed with the model can detect and classify cassava disease in real time, which is equivalent to having a valuable consulting tool. In the future, a cassava disease detection system can be developed based on the classification results of cassava diseases, which can judge the disease categories and provide corresponding management methods. This can greatly improve the planting efficiency of cassava, improve the production efficiency of cassava, achieve scientific and technological progress of agriculture, and promote agriculture into the era of intelligence. Although the model in this study achieved a good success rate in a limited number of cassava diseases, cassava diseases are not limited to these diseases. To improve this, more images can be collected in different cassava planting areas and field conditions, and the model can be more effective in identifying cassava diseases based on field conditions by amplifying the dataset.

Data availability statement

Author contributions

Funding

Acknowledgments

Conflict of interest

Publisher's note

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

References

Aamir, M., Li, Z., Bazai, S., Wagan, R. A., Bhatti, U. A., Nizamani, M. M., et al. (2021). Spatiotemporal change of air-quality patterns in hubei province–a pre-to post-COVID-19 analysis using path analysis and regression. *Atmosphere* 12 (10), 1338. doi: 10.3390/atmos12101338

Bazai, S. U., Jang-Jaccard, J., and Wang, R. (2017). "Anonymizing k-NN classification on MapReduce," in *International conference on mobile networks and management* (Cham: Springer), 364–377.

Bhatti, U. A., Ming-Quan, Z., Qing-Song, H., Ali, S., Hussain, A., Yuhuan, Y., et al. (2021a). Advanced color edge detection using Clifford algebra in satellite images. *IEEE Photonics J.* 13 (2), 1–20. doi: 10.1109/JPHOT.2021.3059703

Bhatti, U. A., Yu, Z., Chanussot, J., Zeeshan, Z., Yuan, L., Luo, W., et al. (2021b). Local similarity-based spatial–spectral fusion hyperspectral image classification with deep CNN and gabor filtering. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3090410

Bhatti, U. A., Nizamani, M. M., and Mengxing, H. (2022a). Climate change threatens pakistan's snow leopards. *Science* 377 (6606), 585–586. doi: 10.1126/science.add9065

Bhatti, U. A., Zeeshan, Z., Nizamani, M. M., Bazai, S., Yu, Z., and Yuan, L. (2022b). Assessing the change of ambient air quality patterns in jiangsu province of China pre-to post-COVID-19. *Chemosphere* 288, 132569. doi: 10.1016/j.chemosphere.2021.132569

Bhatti, U. A., Wu, G., Bazai, S. U., Nawaz, S. A., Baryalai, M., Bhatti, M. A., et al. (2022c). A pre-to post-COVID-19 change of air quality patterns in anhui province using path analysis and regression. *Polish J. Environ. Stud.* 31 (5), 4029–4042. doi: 10.15244/pjoes/148065

Feng, S., Liu, Q., Patel, A., Bazai, S. U., Jin, C. K., Kim, J. S., et al. (2022). Automated pneumothorax triaging in chest X-rays in the new Zealand population using deep-learning algorithms. *J. Med. Imaging Radiat. Oncol.* 66 (8), 1035–1043. doi: 10.1111/1754-9485.13393

Hameed, M., Yang, F., Bazai, S. U., Ghafoor, M. I., Alshehri, A., Khan, I., et al. (2022a). Convolutional autoencoder-based deep learning approach for aerosol emission detection using LiDAR dataset. *J. Sensors* 2022, 17. doi: 10.1155/2022/3690312

Hameed, M., Yang, F., Bazai, S. U., Ghafoor, M. I., Alshehri, A., Khan, I., et al. (2022b). Urbanization detection using LiDAR-based remote sensing images of azad Kashmir using novel 3D CNNs. *J. Sensors* 2022, 9. doi: 10.1155/2022/6430120

Hewitt, C., and Mahmoud, M. (2018). Shape-only features for plant leaf identification. *arXiv preprint arXiv* 1811, 08398. doi: 10.48550/arXiv.1811.08398

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778. doi: 10.48550/arXiv.1512.03385

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. doi: 10.48550/arXiv.1704.04861

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7132–7141. doi: 10.1109/CVPR.2018.00745

Hu, K., Zhang, Z., Niu, X., Zhang, Y., Cao, C., Xiao, F., et al. (2018). Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. *Neurocomputing* 309, 179–191. doi: 10.1016/j.neucom.2018.05.011

Jamil, A., ali Hameed, A., and Bazai, S. U. (2021). Land cover classification using machine learning approaches from high resolution images. *J. Appl. Emerging Sci.* 11 (1), 108. doi: 10.36785/jaes.111501

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

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Commun. ACM.* 60 (6). doi: 10.1145/3065386

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*. 2980–2988. doi: 10.48550/arXiv.1708.02002

Narayanan, K. L., Krishnan, R. S., Robinson, Y. H., Julie, E. G., Vimal, S., Saravanan, V., et al. (2022). Banana plant disease classification using hybrid convolutional neural network. *Comput. Intell. Neurosci.* 2022, 13. doi: 10.1155/2022/9153699

Pujari, D., Yakkundimath, R., and Byadgi, A. S. (2016). SVM and ANN based classification of plant diseases using feature reduction technique. *IJIMAI* 3 (7), 6–14. doi: 10.9781/ijimai.2016.371

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv* 1409, 1556. doi: 10.48550/arXiv.1708.02002

Song, Y., Wang, X., Lei, L., and Xing, Y. (2015). Evaluating dynamic reliability of sensors based on evidence theory and confusion matrix. *Control Decision* 30 (6), 1111–1115. doi: 10.1007/s10489-018-1188-0

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9. doi: 10.1109/CVPR.2015.7298594

Tang, C., Zhu, Q., Wu, W., Huang, W., Hong, C., and Niu, X. (2020). PLANET: improved convolutional neural networks with image enhancement for image classification. *Math. Problems Eng.* 2020, 1–10. doi: 10.1155/2020/1245924

Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning* (PMLR), 6105–6114. doi: 10.48550/arXiv.1905.11946

Turkoglu, M., and Hanbay, D. (2019). Recognition of plant leaves: An approach with hybrid features produced by dividing leaf images into two and four parts. *Appl. Mathematics Comput.* 352, 1–14. doi: 10.1016/j.amc.2019.01.054

Van Hieu, N., and Hien, N. L. H. (2020). Automatic plant image identification of Vietnamese species using deep learning models. *arXiv preprint arXiv* 2005, 02832. doi: 10.14445/22315381/IJETT-V68I4P205S

Wang, H., Shang, S., Wang, D., He, X., Feng, K., and Zhu, H. (2022). Plant disease detection and classification method based on the optimized lightweight YOLOv5 model. *Agriculture* 12 (7), 931. doi: 10.3390/agriculture12070931

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "CBAM: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*. 3–19. doi: 10.1007/978-3-030-01234-2_1

Yu, F., and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv* 1511, 07122. doi: 10.48550/arXiv.1511.07122

Zakzouk, S., Ehab, M., Atef, S., Yousri, R., Tawfik, R. M., and Darweesh, M. S. (2021). "Rice leaf diseases detector based on AlexNet," in *2021 9th international Japan-Africa conference on electronics, communications, and computations (JAC-ECC)* (IEEE), 170–174. doi: 10.1109/JAC-ECC54461.2021.9691435

Zhang, S. (2016). *Research on plant leaf images identification algorithm based on deep learning* (Beijing, China: Beijing Forestry University).

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2921–2929. doi: 10.1109/cvpr.2016.319

Zhu, L., Geng, X., Li, Z., and Liu, C. (2021). Improving yolov5 with attention mechanism for detecting boulders from planetary images. *Remote Sens.* 13 (18), 3776. doi: 10.3390/rs13183776

Zoran, D., Chrzanowski, M., Huang, P. S., Gowal, S., Mott, A., and Kohli, P. (2020). "Towards robust image classification using sequential attention models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9483–9492. doi: 10.48550/arXiv.1912.02184

RADFNet: An infrared and visible image fusion framework based on distributed network

Siling Feng[1], Can Wu[1], Cong Lin[1]* and Mengxing Huang[1,2]*

[1]College of Information and Communication Engineering, Hainan University, Haikou, China, [2]State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China

Introduction: The fusion of infrared and visible images can improve image quality and eliminate the impact of changes in the agricultural working environment on the information perception of intelligent agricultural systems.

Methods: In this paper, a distributed fusion architecture for infrared and visible image fusion is proposed, termed RADFNet, based on residual CNN (RDCNN), edge attention, and multiscale channel attention. The RDCNN-based network realizes image fusion through three channels. It employs a distributed fusion framework to make the most of the fusion output of the previous step. Two channels utilize residual modules with multiscale channel attention to extract the features from infrared and visible images, which are used for fusion in the other channel. Afterward, the extracted features and the fusion results from the previous step are fed to the fusion channel, which can reduce the loss in the target information from the infrared image and the texture information from the visible image. To improve the feature learning effect of the module and information quality in the fused image, we design two loss functions, namely, pixel strength with texture loss and structure similarity with texture loss.

Results and discussion: Extensive experimental results on public datasets demonstrate that our model has superior performance in improving the fusion quality and has achieved comparable results over the state-of-the-art image fusion algorithms in terms of visual effect and quantitative metrics.

KEYWORDS

distributed fusion, multiscale channel attention, edge attention, image enhancement, intelligent agriculture

1 Introduction

Infrared images and visible images are important sensing information for intelligent agricultural systems. The key to intelligent agricultural systems is to utilize perceptual data for intelligent analysis and decision-making. The infrared imaging technology with anti-interference solid ability uses the radiation energy released by the target so it can penetrate smoke, fog, rain, snow, etc., in the environment. However, the visible light sensor uses light

reflectivity to image with much spectral information and high-resolution characteristics. As the application range in intelligent agricultural equipment gradually broadens and the perceived information environment is usually changeable, a single image imaging technology cannot sufficiently perceive the environmental information. It results in the inability of intelligent agricultural equipment to perceive enough information, which leads to the failure of intelligent agricultural systems to work regularly. Therefore, it is of great significance to study the complementary use of infrared and visible image imaging technology to enhance the information perception ability of intelligent agricultural equipment (Aamir et al., 2021).

High-quality enhanced images can be obtained by fusing infrared and visible images to improve the information perception ability in intelligent agricultural equipment and meet various subsequent visual tasks for intelligent agricultural systems. As a branch of information fusion, image fusion has played an essential role in computer vision since it can generate more informative images for high-level vision tasks such as recognition (Basak et al., 2022), detection (Wieczorek et al., 2022), tracking (Bhatti et al., 2022d; Yan and Woźniak, 2022), and surveillance (Chen et al., 2021; Chen et al., 2022b). Significantly, infrared and visible image fusion is a considerable problem and has striking advantages. It is a task that aims to integrate salient features extracted from source images into a single image by appropriate methods (Li et al., 2017). Generally, visible images contain texture information with high spatial resolution and often lose effectiveness under dark or extreme environmental conditions. On the contrary, infrared images can highlight thermal targets in low light or severe weather and contain little texture information because of their low spatial resolution. Infrared and visible image fusion can integrate the complementary virtues from infrared and visible images into synthetic images, which not only conform to human visual perception but also adapt to the application in various vision systems (Bhatti et al., 2022a; Bhatti et al., 2022b).

According to the abstract degree in image information, image fusion is divided into three levels: pixel level, feature level, and decision level (Ma et al., 2019a). In this work, we mainly study pixel-level image fusion methods because they can retain the information from the source image to the maximum extent. In the past decades, scholars have proposed numerous infrared and visible image fusion techniques. These approaches can be broadly classified into two categories: traditional and deep learning-based methods (Ma et al., 2019a). Most traditional infrared and visible image fusion algorithms belonging to pixel-level fusion directly perform mathematical operations on the image pairs after image registration, which have achieved good performance. However, infrared and visible image fusion methods based on deep learning have emerged with tremendous potential and even better performance in recent years.

The traditional methods, in general, cover five approaches: multi-scale transform methods (MST) (Zhu et al., 2018), sparse representation methods (SR) (Cui et al., 2015; Zhang et al., 2018), saliency methods, subspace methods, and other methods (Gangapure et al., 2018). In general, MST-based methods first decompose the source images into multiple scales, and then the multi-scale features are fused using the appropriate fusion rule. Finally, an inverse operation is performed to reconstruct the fused image. The MST

based methods usually adopt Laplacian pyramid transform (LP) (Bulanon et al., 2009), wavelet transform (Wavelet) (Mallat, 1989), nonsubsampled contourlet transform (NSCT) (Da Cunha et al., 2006), edge-preserving filter (EPF) (Farbman et al., 2008), curvelet transform (CVT) (Nencini et al., 2007), and multi-resolution singular value decomposition (MSVD) (Naidu, 2011). Sparse representation methods (SR) generally comprise four steps (Ma et al., 2019a): First, a sliding window strategy is adopted to decompose the source image into several overlapping patches. Then a learned over-complete dictionary is used for sparse coding on each patch to obtain the sparse representation coefficients. Thirdly, a reasonable fusion strategy is designed to fuse sparse representation coefficients. Finally, the learned over-complete dictionary produces a marked effect in reconstructing the fused image using the fused coefficients. Among them, the construction of the over-complete dictionary is key in SR (Ma et al., 2019a). The saliency-based methods can highlight regional activity and significance (Meng et al., 2017; Zhang et al., 2017). The subspace-based methods, including the principal component analysis (Bavirisetti et al., 2017), independent component analysis (Mitianoudis et al., 2013), and non-negative matrix factorization (Kong et al., 2014) can remove the redundant information existing in most natural images by converting high dimensional input images into low dimensional spaces or subspaces. Although the existing traditional fusion methods have indicated great performance, these methods require the highly manual design in decomposition and fusion strategies. Their application is subject to unpredictable constraints in some tasks, and their performance deteriorates when the source images are complex due to the degradation of representation (Chen et al., 2022a).

In the past several years, deep learning has been widely applied in infrared and visible image fusion to solve the shortcomings in traditional fusion methods. The application of deep learning-based methods for infrared and visible image fusion mainly reflects in convolutional neural network CNN-based network frameworks, such as convolutional sparse representation (CSR) and generative adversarial network (GAN). The CNN-based fusion frameworks for infrared and visible image fusion are divided into two categories: the depth extraction for image features and the construction for fusion networks. In depth feature extraction, VGG-19 (Ren et al., 2018), ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152 (Szegedy et al., 2017) have been proposed, among which VGG-19 and ResNet152 are commonly applied. The depth of ResNet152 is deeper than that of VGG-19, and deepening network depth improves the depth features in the image. Nevertheless, the more convolution layer parameter maps cause the problems in increasing the number of parameters, the amount of calculation, and the high requirement for computing hardware. The CSR-based methods generally combine PCNN, wavelet transform, and NSCT to construct a fusion network structure, which has been widely used in infrared and visible image fusion. They can effectively represent the salient features in the source images. However, the local modeling approach adopted by image fusion methods based on sparse representation is prone to lead to two major defects: loss of contextualized information and low tolerance of fault matching. The GAN-based fusion algorithms adopt the CNN network structure as the framework with strong feature extraction ability, significantly improve the fusion quality, and use the confrontation

between the source image and the generated image to realize the supervision in the source image on the learning parameters. Ma et al. introduced the GAN in the infrared and visible image fusion task for the first time, namely FusionGAN (Ma et al., 2019b), and then more GAN-based fusion frameworks are proposed (Ma et al., 2020; Li et al., 2021b; Ma et al., 2021). Nevertheless, they are limited by the size of the convolution kernel and the depth of the network, ignoring the correlation between the feature map channels.

Although a variety of networks to improve the performance in image fusion have been proposed by many scholars. The CNN-based network frameworks, such as convolutional sparse representation (CSR), generative adversarial network (GAN), and other many network architectures are applied in infrared and visible image fusion. However, the CNN-based fusion frameworks for infrared and visible image fusion are divided into two categories: the depth extraction for image features and the construction of fusion networks. The extraction for depth features requires a deeper network structure, resulting in weak interpretability, extensive computation, and other problems. The construction of the fusion network is also complex and difficult to control. Although many modelsare superficially similar to RADFNet, they have not abandoned these two categories. To get rid of the dilemma in these two kinds of fusion categories, the RADFNet employs a distributed fusion framework to make the most of the fusion output from the previous step. Two channels utilize residual modules with multiscale channel attention to extract the features from infrared and visible images, which are used for fusion in the other channel. Because it adopts distributed fusion, the fusion network does not entirely rely on the extraction in deep features, and the fusion network is simple to construct, showing strong robustness. The RADFNet solves the limitations from most current fusion networks and shows strong adaptability. The main contributions of our work are summarized as follows:

(A) A distributed fusion framework based on residual CNN (RDCNN) for infrared and visible image fusion is proposed in this paper. The distributed fusion framework is distinct from the existing fusion framework in infrared and visible image fusion. It adopts three channels to realize image fusion, wherein two channels are applied to feature extraction and the other channel realizes feature fusion.

(B) To obtain coarse-to-fine features and compensate edge information for fused images, the attention mechanism is discussed. In this way, the fused images retain more prominent information and lose less edge information from source images.

(C) Two loss functions, including the pixel intensity with texture loss and the structural similarity (SSIM) with texture loss, are designed to train the RADFNet. Through experiments, it is found that networks trained by the two loss functions have their own advantages.

(D) Extensive experiments are conducted on public infrared and visible image fusion datasets. Compared with existing state-of-the-art fusion methods, our fusion framework has a promising even better performance in accordance with visual effect and quantitative metrics. In addition, we perform ablation experiments to verify the function in the corresponding module. Last but not least, unregistered source image pairs are fed into the proposed network, emerging the robustness of the proposed framework.

2 Materials and methods

2.1 Related works

2.1.1 Distributed fusion architecture

Distributed fusion architecture is a classical and typical structure in multi-sensor fusion due to its high speed and reliability (Sun et al., 2017). In distributed fusion, the measurement results of each sensor are processed to obtain local estimates and error covariance. Then the processing results are sent to the fusion node to conflate them into global state estimation and the estimated error covariance (Wu et al., 2021). Figure 1 shows a distributed model for the fusion in radar and infrared sensors (Yang et al., 2016). For single target tracking, radar and infrared sensors track the target respectively and generate dependent target trajectories in their local information processing center, then send the local trajectory information to the fusion center for data fusion.

2.1.2 Residual network

In some tasks, deeper neural networks can extract higher-level features and perform excellently. However, too deep networks may cause the notorious problem of vanishing or exploding gradients and degrade the accuracy. To solve these problems, He et al. proposed a residual network composed of a series of residual blocks (He et al.,
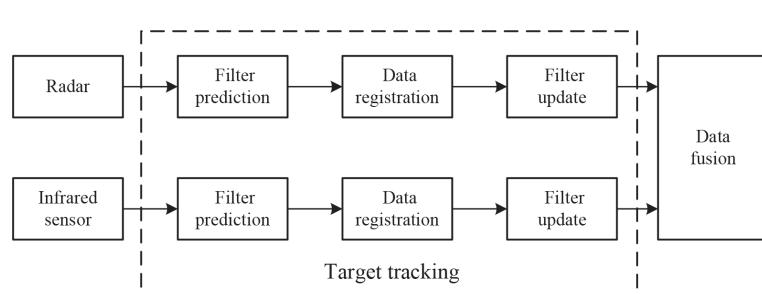

FIGURE 1
A distributed fusion model for radar and infrared sensors.

2016a). Figure 2A shows the original residual module, which can be expressed as (He et al., 2016b).

$$y_l = h(x_l) + \mathcal{F}(x_l, W_l)$$
$$x_{l+1} = f(y_l) \tag{1}$$

where x_l and x_{l+1} are the input and output in the l-th layer, and $\mathcal{F}$ is the residual function. f is a ReLU (Nair and Hinton, 2010) function. The residual block contains two parts: identity mapping and residual mapping. The left part of Figure 2A is the identity mapping, and the right part of Figure 2A is the residual part expressed as $\mathcal{F}(x_l, W_l)$, which usually contains 2 or 3 convolutional layers. In many cases, the dimensions of input x_l and output x_{l+1} are discrepant, so it is necessary to employ a 1×1 convolution operation to maintain the dimension in input and output consistent, whose schematic diagram is shown in Figure 2B, which can be expressed as (He et al., 2016b).

$$x_{l+1} = h(x_l) + \mathcal{F}(x_l, W_l)$$
$$h(x_l) = W_l' x_l \tag{2}$$

where h(x_l) is the identity skip connection and W_l' is the 1×1 convolution kernel.

The residual network can be formulated as (He et al., 2016b)

$$x_L = x_l + \sum_{i=l}^{L-1} \mathcal{F}(x_i, W_i) \tag{3}$$

for any deeper block L and any shallower block l. The formula 3 indicates the feature x_L in any deeper residual block L which can be represented as the feature x_l in any shallower block l add the residual function, which leads to nice backward propagation properties that the gradient of layers will not vanish even when the weights are arbitrarily small (He et al., 2016b). Moreover, experiments with the various usages of activation function were carried out in (He et al., 2016b). The order of the activation function in the network will affect the performance of the residual network. The structure of the improved residual unit shown in Figure 2C has the best performance. In this structure, the batch normalization (BN) and ReLU activation function are placed before the convolution layer, and the activation function after addition is moved to the residual part.

2.1.3 Attention mechanism in deep learning

Attention mechanism can be traced to the last century, which was mostly applied to machine translation tasks. It has become an essential concept in artificial intelligence because it conforms to some laws of human cognition and can improve the interpretability of neural networks. Therefore, the attention mechanism is widely applied, such as natural language processing, speech recognition and computer vision (Mnih et al., 2014; Vaswani et al., 2017; Bhatti et al., 2022c). In the computer vision domain, many researchers have studied attention mechanism and proposed corresponding methods to acquire nice performance. A residual attention network built by stacking attention modules is proposed in (Wang et al., 2017) which are designed to generate attention-aware features, achieving outstanding recognition performance. A novel architecture unit termed the "Squeeze-and-Excitatio"(SE) block that adaptively recalibrates the channel feature strength by explicitly modelling the interdependence between channels is introduced in (Hu et al., 2020). The structure of SE block is shown in Figure 3, where U is a feature map with the size of W×H×C, ⊗ and refers to channel-wise multiplication, so X and U have the same size. Moreover, edge-guided attention mechanisms which can produce visually appealing images also attract the attention of many researchers (Bhatti et al., 2021). Zhao et al. (Zhao et al., 2019a) propose an edge guidance network (EGNet) which solves the problems of rough boundary in object detection through the complementarity of the object and salient edge information.

2.2 Methods

2.2.1 Overall framework

Enlighted by the advantages of distributed structure and the residual module, we propose a novel distributed fusion architecture for infrared and visible images based on the residual module and attention of edge and multiscale channel, RADFNet. The RADFNet is an end-to-end fusion network, the overall structure of which is shown in Figure 4. It contains four parts: the feature extraction for the visible image, the feature extraction for the infrared image, the fusion for
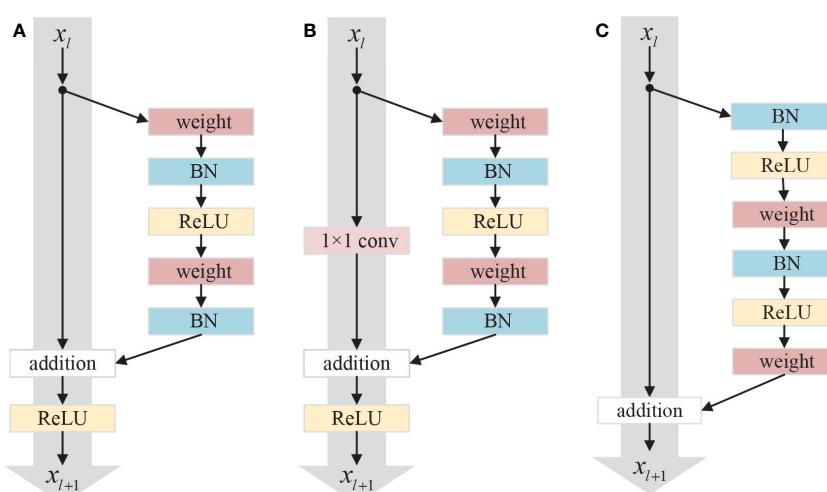

FIGURE 2
Three different residual units: **(A)** Original residual unit; **(B)** Conv residual unit; **(C)** Improved residual unit.
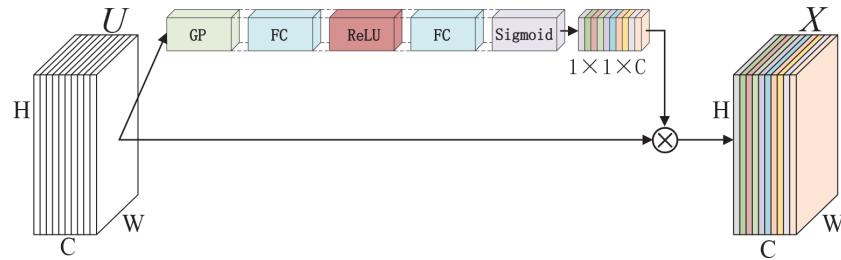
FIGURE 3
A Squeeze-and-Excitation block, where GP means global average pooling, FC refers to fully-connected layers, ReLU refers to the ReLU function, and Sigmoid refers to the sigmoid function.

features, and the compensation for edge information. The infrared and visible image fusion process is formulated as follows.

The visible image features extraction branch can be formulated as

$$V_i = ATT_i(VR_i(V_{i-1})) \oplus E_{V_i} \quad i = 1, 2, 3, 4 \qquad (4)$$

$$E_{V_i} = EAT_i(E_{V_{i-1}}) \quad i = 1, 2, 3, 4 \qquad (5)$$

where V_0, the visible image input in the architecture, is the V in the Figure 4. V_i is the representation of V_0 after the residual module, multiscale channel attention and compensation of edge information. VR_i means the residual module acting on the V_{i-1} and ATT_i is the multiscale channel attention module designed to obtain coarse-to-fine features from the outcome of VR_{i-1}. Vi represents the features in different levels of V_0 with different scales, wherein V_i has a higher level than V_{i-1}. E_{Vi} is the edge information feature map obtained by EAT_i with input $E_{V_{i-1}}$ configured to compensate for the edge information of the feature map achieved by residual module and multiscale channel attention module. ⊕ refers to the maximum value in the homologous channel and position in the feature map. The features in the visible image with separate scales are extracted through the above steps. Then, they are fed into the fusion channel to fuse at each layer, which can fully utilize the multi-scale information from perceptible images. In this method, more texture information with high spatial resolution retains, which can enhance the quality of the fused image.

The infrared image feature extraction branch can be formulated as

$$I_i = ATT_i\left(\underset{i}{\mathrm{IR}}(I_{i-1})\right) \oplus E_{I_i} \quad i = 1, 2, 3, 4 \qquad (6)$$

$$E_{I_i} = EAT_i(E_{I_{i-1}}) \quad i = 1, 2, 3, 4 \qquad (7)$$

where I_0, the infrared image input in the architecture, is the I in Figure 4. I_i is the representation of I_0 after the residual module,

multiscale channel attention and compensation of edge information. IR_i means the residual module acting on the I_{i-1} and ATT_i is the multiscale channel attention module designed to obtain coarse-to-fine features from the outcome of IR_{i-1}. I_i represents the features in different levels of I_0 with different scales, wherein I_i has ahigher level than I_{i-1}. E_{I_i} is the edge information feature map obtained by EAT_i with input $E_{I_{i-1}}$ configured to compensate for the edge information of the feature map achieved by residual module and multiscale channel attention module. ⊕ refers to the maximum value in the homologous channel and position in the feature map. The features in the infrared image with distinct scales are extracted through the above steps. Then they are constituted into the fusion channel to fuse at each layer, which can fully utilize the multi-scale information from infrared images. As a result, rich target information is used for highlighting the target in the fused image.

The channel of feature fusion can be defined as

$$FUI_i = \begin{cases} F_i(V_i, I_i) & i = 0 \\ F_i(V_i, I_i, FUI_{i-1}) & i = 1, 2, 3, 4 \end{cases} \qquad (8)$$

where V_0 and I_0, which are visible image and infrared image inputs in the fusion architecture, are the V and I in Figure 4 respectively. FUI_1, FUI_2, FUI_3, and FUI_4 are the fusion results with different level features using corresponding rules. F_i refers to the fusion rule of the relevant layer features. FUI_i is the fusion result of the i-th extracted features V_i, I_i and the different scales from previous fusion result FUI_{i-1}. It realizes the layer-by-layer fusion so it can make the best use of the information from multisource images and then improve the quality of the fused image.

2.2.2 Network structure

The infrared and visible image fusion model RADFNet set out in the present paper is constituted of three channels. The RADFNet
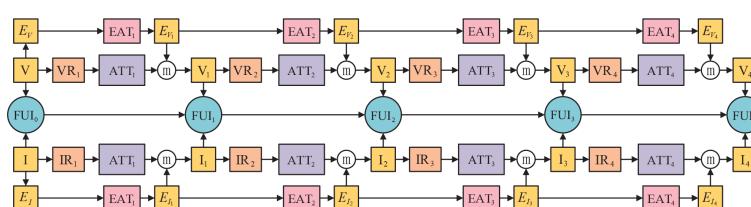
FIGURE 4
The overall structure for infrared and visible image fusion.

structure is exhibited in Figure 5. RADFNet contains four parts: the features extraction branch of the visible image and infrared image, the features fusion branch, and the edge attention module compensating edge information for the extracted features. The left and right branches in Figure 5 are intended to extract the features in visible and infrared images respectively. The middle branch fuses the features extracted by the two branches with the results from the previous step layer by layer, and the last layer generates the fused image. For a convolutional layer, '$k \times k,(in,out)$' means the kernel size is $k \times k$, the input channel is *in* and the output channel is *out*. In the network, BN indicates batch normalization that is utilized to speed up the training and make the training more stable, and ReLU denotes the linear rectification function.

The RADFNet adopts four-layers network structure. The VR_{1-4} and IR_{1-4} are the residual networks which extract image features.

Because the residual network has the advantages of mitigating gradient disappearance or gradient explosion and protecting the information integrity, the networks we designed can extract meaningful features and ensure the information integrity simultaneously. Besides, $ATTi$ processes the features extracted by residual block VRi or IRi to obtain coarse-to-fine features. $EATi$ acquires the edge information and then compensates edge information for the extracted feature map. The ⊕ refers to the operation for achieving the maximum value in the homologous channel and the homologous position in the feature map. The $FUPi$ generates $FUIi$ by fusing features extracted by the other two branches with the FUI_{i-1} generated by FUP_{i-1} when i is not 1. When i is 1, the concatenated infrared and visible image is fed into the $FUP1$ to generate FUI_1. The ⊕ is the concatenation operation in channel-wise, and the 1×1 244 convolution layer in the last fusion layer constructs fusion images.
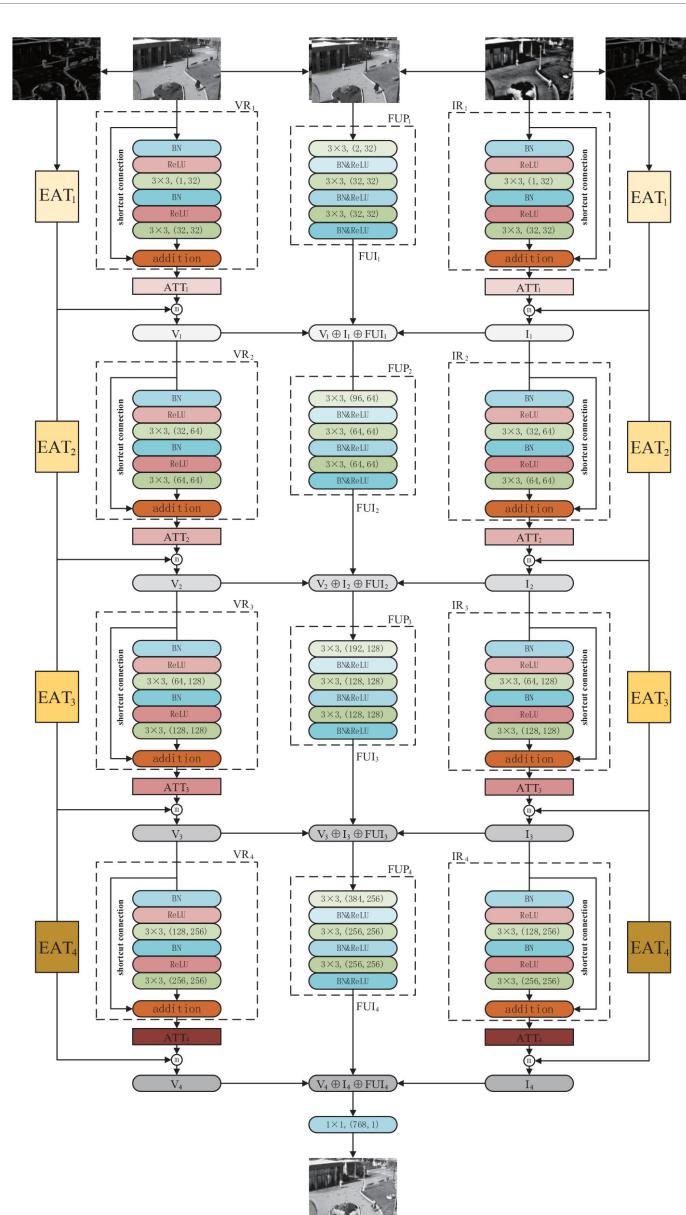
FIGURE 5
The structure of RADFNet. 'ATT_{1-4}' denote the multiscale channel attention module and 'EAT_{1-4}' denote the edge attention module. '$3 \times 3,(1, 32)$' means the kernel size is 3×3, input channel is 1 and output channel is 32 in a convolutional layer.

2.2.3 Multiscale channel attention network

In the process of infrared and visible image fusion, image feature extraction is exceptionally significant. However, in practical situations, numerous detailed information loses in the process of feature extraction. Inspired by SENet (Hu et al., 2020), the multiscale channel attention network is proposed to process the features extracted by the residual network to obtain the coarse-to-fine features, which can retain more detailed information in the feature map. As shown in Figure 6, the structure enclosed by the dotted line is the multiscale channel attention module. The features which lose a lot of details extracted by VRi or IRi are used as input in ATTi. Then, the 1×1, 2×2, and 4×4 average pooling operations are performed to generate multiscale features which contain more necessary spatial information. Moreover, the channel attention mechanism is utilized to enhance channel correlation information between features. The multiscale channel attention network is trained to learn the weight $W_{t_i}^k$ for the k-th feature $f_{t_i}^k$ of the t-th pooling scale in the ATTi which can be formulated as

$$W_{t_i}^k = \sigma(w_2 \delta(w_1 G(z))) \tag{9}$$

$$G(z) = \frac{\sum_{x,y} f_{t_i}^k(x,y)}{H \times W} \tag{10}$$

where $G(z)$ denotes the global average pooling operation. $\sum_{x,y} f_{t_i}^k(x,y)$ means the sum of the k-th feature with the t-th pooling scale in ATTi. (x, y) refers to the position in feature map, and H,W means the height and width of the feature map. δ refers to the ReLU function, $w_1 \in \mathbb{R}^{k \times k}$ and $w_2 \in \mathbb{R}^{k \times k}$, σ denotes the sigmoid function. Then the channel-wise multiplication is implemented between $W_{t_i}^k$ and the up-sampled features which can be expressed as $UP(f_{t_i}^k)$, ensuring the multiscale features have the same size as the input. Based on this, the reweighted features are obtained and then the attention map can be achieved as follows:

$$F_i = \pi\left(W_{1_i}^k * UP\left(f_{1_i}^k\right)\right) \textcircled{m} \pi\left(W_{2_i}^k * UP\left(f_{2_i}^k\right)\right) \textcircled{m} \pi\left(W_{3_i}^k * UP\left(f_{3_i}^k\right)\right) \tag{11}$$

where π denotes the instance normalization (Ulyanov et al., 2016) and $\textcircled{m}$ refers to the operation for acquiring the maximum value in the homologous channel and position in the feature map. Through the above method, the coarse-to-fine attention map Fi is obtained. The attention

map not only emphasizes more critical features and neglects secondary ones but also reserves more necessarily detailed information.

2.2.4 Edge attention module

Generally, the edge information of an image refers to the sudden change in local grayscale value, color component and texture structure. The edge information from images which is helpful to distinguish objects, can effectively attract attention of people due to human visual characteristics. Enlightened by previous work, we utilize an edge feature map extraction model from the shallower to deeper to obtain the enhanced edge maps, which are designed to compensate for textural information for the fused image.

For the sake of acquiring the edge information used to compensate fused images, we obtain the gradient map from the source images. The process of obtaining the gradient maps ∇g by inputting a gray-scale image f with the size $h \times w$ is defined as

$$\nabla g = \sum_{x=1,y=1}^{x=h-1,y=w-1} \sqrt{\left(\nabla g^h(x,y)\right)^2 + \left(\nabla g^w(x,y)\right)^2} \tag{12}$$

$$\begin{aligned} \nabla g^h(x,y) &= f(x,y) - f(x+1,y) \\ \nabla g^w(x,y) &= f(x,y) - f(x,y+1) \end{aligned} \tag{13}$$

where $f(x, y)$ means the pixel at position (x, y). Moreover, we perform the enhanced operation to obtain the more obvious gradient information:

$$G = \max_{y \in W} \max_{x \in H}\left(\nabla g(x+1,y+1), \nabla g(x,y)\right). \tag{14}$$

where $H=\{1,…,h-1\}$ and $W=\{1,…,w-1\}$. The (x, y) represent the position at the gradient map. Through the above steps, we get the gradient image G with the abundant enhanced edge information.

Subsequently, we feed the gradient images from infrared and visible images into the edge attention module to generate edge attention feature maps with enhanced edge information. Then, the feature maps will be entered into the extraction branch to compensate edge information for the extracted features by IRi or VRi. The structure diagram of the edge attention module is shown in Figure 7. The edge attention module generates E_{Vi} and E_{Ii} layer by layer, which is then used to compensate edge information for the
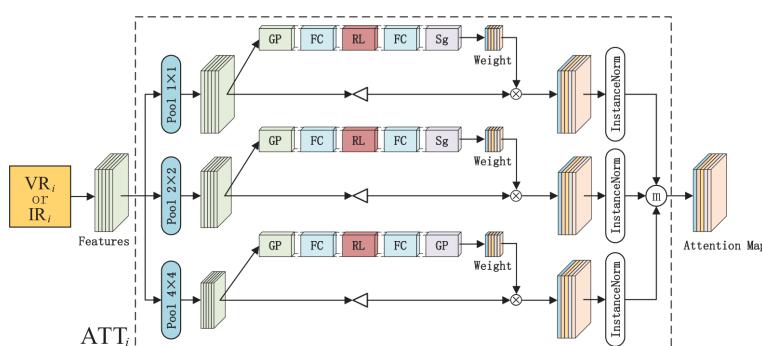

FIGURE 6
The multiscale channel attention network. The features extracted by residual network(VRi or IRi) are fed into the ATTi to generate attention map. GP, FC, RL, SG denote the global average pooling operation, fully connected layer, ReLUfunction and sigmoid function respectively. ◁ stands for the up-sample operation and ⊗ denotes the element-wise multiplication.

3 Experimental results and analysis

3.1 Experimental configurations

2.2.5 Loss function

For infrared and visible image fusion, it is difficult to provide the ground truth of fused images for networks to train a model. However, the requirement to retain salient target information in the infrared image and the texture information in the visible image is determined. Inspired by this requirement, the loss function we employ is as follows:

$$L_F = L_{\text{pixel}} + a L_{\text{texture}} \qquad (15)$$

where the L_{pixel} constrains the fused image to contain more target information from the image pair facilitating target tracking and the $L_{texture}$ forces the fused images to contain more texture details which can effectively improve the identification of objects in images.

Specifically, the exact definition of L_{pixel} is expressed as follows:

$$L_{\text{pixel}} = \frac{1}{m} \sum_{j=1}^{m} \| I_f^j - \max\left(I_{ir}^j, I_{vis}^j\right) \|^2 \qquad (16)$$

where m is the batch size that is the number of training samples used in each iteration. The I_f means the fused image with the input image pair $\{I_{ir}, I_{vis}\}$, and the max $(\cdot)$ denotes the element-wise maximum selection. Through the maximum selection strategy, the fused images have the prominent target information.

Moreover, we hope the fused images contain significant target information and simultaneously preserve great textural details from source images. However, the L_{pixel} has very limited constraints on textural details. Therefore, the $L_{texture}$ is introduced to force the fused image to retain more textural information and the $L_{texture}$ is defined as:

$$L_{\text{texture}} = \frac{1}{m} \sum_{j=1}^{m} \| \left|\nabla I_f^j\right| - \max\left(\left|\nabla I_{ir}^j\right|, \left|\nabla I_{vis}^j\right|\right) \|^2 \qquad (17)$$

where the m is the batch size, the I_f means the fused image with the input image pair $\{I_{ir}, I_{vis}\}$, and the max $(\cdot)$ denotes the element-wise maximum selection. The ∇ indicates the Sobel gradient operator and the $|\cdot|$ means the absolute operation. The element-wise maximum selection strategy can make the fused images obtain the most significant edge textural information.

To evaluate the proposed fusion algorithm in many aspects, we conduct extensively qualitative and quantitative experiments on the RoadScene (Xu et al., 2020) dataset. We evaluate the performance of our method by making a comparison with six state-of-the-art approaches, including two Nest-based methods, i.e., NestFuse (Li et al., 2020) and RFN-Nest (Li et al., 2021a), and four CNN-based methods: DenseFuse (Li and Wu, 2018), IFCNN (Zhang et al., 2020), U2Fusion (Xu et al., 2022), and SDNet (Zhang and Ma, 2021). The subjective visual perception system is vulnerable to human factors, such as personal emotion and visual environment, and the fused images using different approaches resemble somewhat. Therefore, there are six evaluation statistical metrics which are selected to quantify the evaluation, including mutual information(MI) (Qu et al., 2002), entropy(EN) (Roberts et al., 2008), visual information fidelity(VIF) (Han et al., 2013), stand deviation (SD), spatial frequency(SF) (Eskicioglu and Fisher, 1995) and average gradient (AG) (Zhao et al., 2019b). MI quantifies the amount of information obtained from the source image by the fused image, and EN assesses the amount of information contained in the fused image based on information theory. VIF mainly computes information fidelity in a fused image, which is in line with human visual perception. SD reflects the contrast of an image based on statical concepts, a larger SD value indicates a higher contrast distribution in an image, and the image carries more information. SF reflects the change rate of image gray scale. AG can measure the fused image clarity, which can be considered that the greater AG, the better the image clarity and the better the fused image quality. EN, SF and SD are reference-free metrics. Moreover, a fusion method with larger MI, EN, VIF, SD, SF, and AG represents better performance.

3.2 Details of implementation

In the training process of the RADFNet model, we use images from the OSU (Davis and Sharma, 2007) dataset to construct the training dataset. Due to different imaging sensors, the image pairs in the OSU dataset are not strictly registered resulting in black edges in
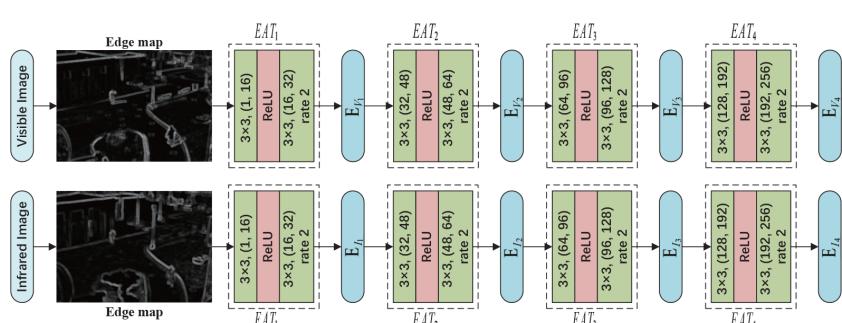
**FIGURE 7**
The architecture of edge attention module. The EAT_{1-4} are designed to generate shallow to deep edge feature maps $E_{V_{1-4}}$ or $E_{I_{1-4}}$. For convolution layer, the 'k×k,(in,out)' means that the convolution kernel size is k×k , the input channel is in and output channel is out. In addition, the 'rate 2' denotes the dilated convolution operator with a dilation rate of two.

infrared images. Therefore, we crop both infrared and visible images at the same size 280 × 200. Based on the above operations, we can get 8,544 image pairs. It is worth nothing that the visible images in the OSU dataset are color images, but the infrared images are grayscale. To make the number of channels with the input image pair the same, we perform the process that converts the visible images to grayscale images in advance. Moreover, all images are normalized to [0,1] before being fed into the network to accelerate model convergence. The hyper-parameter of the loss is set as a = 10. Adam optimizer (Kingma and Ba, 2015) with β1 of 0.9, β2 of 0.999, epsilon of 10^{-8}, weight decay of 0, the initial learning rate of 0.001 is used to optimize our fusion model with the guidance of loss function L_F. All experiments are conducted on the Quadro RTX6000 GPU and 2.90 GHz Intel(R) Xeon(R) Gold 6226R CPU.

The RoadScene dataset contains color visible images, but we employ the input grayscale images to train the proposed network. To get better visuals in the test phase, we adopt the strategy (Prabhakar et al., 2017) to process color images instead of converting the input color images to grayscale images. Precisely, we first convert the color image to the YCbCr color space, then the infrared image and the Y channel of visible image are entered into the RADFNet. Finally, the fusion result is concatenated with Cb andCr channels from visible image along channel-wise and then converted into the RGB color image. The RGB color image is the result of the proposed network.

3.3 Results analysis on RoadScene datasets

To fully evaluate the performance of the RADFNet, we compare the RADFNet with the other six methods on the Roadscene dataset. The Roadscene dataset mainly contains road scenes, including pedestrians and cars, in the daytime and at night. We select two images in the daytime and two in the nighttime for evaluation subjectively so as to exhibit some intuitive fused images on the fusion performance. The fused images of the proposed RADFNet and the other six methods are presented in Figure 8. In the daytime scenes, the fused images with exceptional visual quality have rich texture information from visible images and enhanced prominent target information from infrared images. In the first column images in Figure 8, RADFNet makes the pedestrians in the image have the most incredible vigorous light intensity. The fused images of U2fusion and SDnet show they tend to darken the entire color of the images. For example, the color of the sky is darker than the fusion images with other methods. In the second column, all six methods enhance the pedestrian. Still, all other methods, except the RADFNet, dim the streetlamp to a certain extent, thus losing information. Moreover, the fusion image of the proposed approach has more obvious color contrast and texture details, so the buildings in our fused image have a richer structure sense than the fused images with other methods. In the nighttime scenes, the ability of both infrared images and visible images to provide information is limited. Therefore, sufficiently retaining meaningful data from the source images is challenging. In the third column, all fusion methods inevitably integrate useless information into the fused image, which

degrades the visual quality of the image. Regardless, the proposed approach best protects the information from the visible image while using the meaningful information from the infrared image to enhance the target information. In the last column, compared with other fused images, the fused image in the proposed method failsto remove the halo on the streetlamp altogether. Nonetheless, the signs on the road are most conspicuous in the fusion image, while signs on the street in other images even tend to disappear. In a word, the proposed method can efficiently utilize the information of the infrared and visible images to generate high-quality fused images.

To avoid human factors and other factors affecting the subjective evaluation. We conduct quantitative assessments with the six approaches and the proposed method. The results of six metrics on the Roadscene dataset, which contains 221 image pairs, are shown in Figure 9. It can be noted that our results achieve better performance on six metrics. The best MI means that our method transfers the most information from the source image to the fused image and the best EN represents the fused image thatcontains the most information. The proposed method represents the best on VIF, which indicates our fused image gets a better human visual perception effect. The best SF and AG suggest that the proposed approach generates the clearest image with remarkable quality. In addition, our RADFNet displays the best SD, illustrating our fused images have the highest contrast. Combined with subjective and quantitative evaluation results, these results prove that RADFNet can convert more meaningful information from infrared and visible images to fused images while ensuring the best quality.

3.4 Ablation experiment

To verify the effectiveness of the edge attention module, we conduct ablation experiments. We employ edge attention and ignore edge attention to create two models, then the same image pair is used as input to test the difference between the two models, and the visual results are presented in Figure 10. The red and green box parts are magnified for a more intuitive comparison. In the first row, the telegraph pole in the red box with edge attention has a clearer texture, while that without edge attention even becomes blurred. In addition, the leaves with edge attention in the green box also have more precise texture details than that not using edge attention. The words in the red box of the images in the second row are difficult to identify because of the blurred source image. In contrast, words in the fused image using edge attention are more beneficial to observe than that in the image not using edge attention because the edge attention module compensates for the edge information for the fused image.

In addition, to comprehensively evaluate the impact of edge attention in fused images, we make quantitative evaluations for the four images in Figure 10, and the result is listed in Table 1. It is noted that only the fused images with edge attention have a slightly lower metric SD than that without edge attention. The fused images with edge attention are higher for the other five metrics, i.e., EN, SF, SD, MI, VIF, and AG in both Street and House images. The results show that the generated edge information from the edge attention module compensating for the fused image can improve the image quality effectively.

FIGURE 8
The visual results comparison with different methods on the Roadscene dataset.

3.5 Discussion on loss function

For the sake of comprehensively considering the improvement in model training on fused image quality, we design another loss function L_{FS}, which can be defined as follows:

$$L_{FS} = \beta L_{SSIM} + L_{texture} \tag{18}$$

where $L_{texture}$ is represented by Equation 17, the value of β is 5, and the L_{SSIM} is the structural similarity (SSIM) loss, which can be expressed as

$$L_{SSIM} = 1 - (w \cdot \text{SSIM}(F, I) + (1 - w) \cdot \text{SSIM}(F, V)) \tag{19}$$

where the $SSIM(\cdot)$ means the structural similarity (Wang et al., 2004). F denotes the output result from the proposed model. V and I refer to the homologous input visible and infrared images

respectively. In addition, to balance the structural similarity loss between the fused image and infrared and visible image, the weight w is taken as 0.5.

The loss functions L_F and L_{FS} are used to train the proposed network respectively, and the results are exhibited in Figure 11. In the first row, the zebra crossing in the green box of fused image output after the network trained with L_F is more prominent than that trained with L_{FS}. However, the halo on the streetlamp in the red box in the image output by the network trained by L_F is not completely removed. In the second row, it can be seen that no matter the definition of the whole image or the details, the network output image using L_F training is better. In a word, the output image from the network trained by L_F can highlight more important information in the nighttime scenes. But that trained by L_{FS} can essentially eliminate the halo in the image. In the daytime scenes, the quality of the output

FIGURE 9
The quantization results of six metrics on the 221 image pairs from the Roadscene dataset. The abscissa x refers to the number of image pairs and the ordinate y refers to the metric value.

No edge attention   With edge attention

FIGURE 10
The results of ablation study about the influence of edge attention module in two image pairs from the Roadscene dataset.

TABLE 1 **The quantitative results on the four images shown in** Figure 10.

		EN	SF	SD	MI	VIF	AG
Street	Edge	7.514	0.077	10.492	3.890	0.730	7.763
	No-Edge	7.533	0.061	10.721	2.376	0.590	5.905
House	Edge	7.586	0.072	10.334	3.946	0.937	6.846
	No-Edge	7.573	0.056	10.605	2.686	0.709	5.427

The Street means the first row images and House denotes the second row images. Edge and No-Edge refer to edge attention and no edge attention during image fusion, respectively.
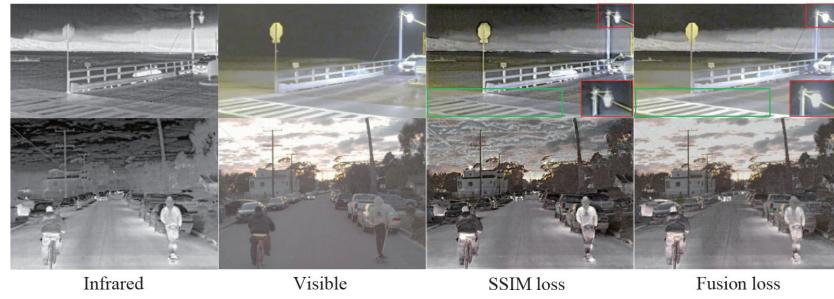
FIGURE 11
The results of RADFNet trained by SSIM loss L_{FS} and Fusion loss L_F.

images from the network trained by L_F is better in both overall and detail. Therefore, we choose fusion loss L_F as the training loss function in our experimental test.

To set the best optimal coefficients for the proposed method, the parameter a is set as 1, 10, 50 and 100. The epoch and batch size are 4 and 4, respectively. One a is needed to choose for the image fusion task based on the test images. Six metrics are employed to evaluate the performance of RADFNet with different a. The values are shown in Table 2. The best values are indicated in red and the second-best values are denoted in blue. It is worth nothing that three of the six metrics are best when a=1. However, the metrics MI and VIF are unstable. When a=10, the values of all metrics are considerable and stable, which indicates the proposed network can achieve better fusion performance than other values of a. So, a is set as 10 in experiments.

3.6 Fusion of unregistered image pairs

In general, it is difficult to obtain the source image pairs that have been strictly registered for image fusion because the imaging characteristics of different sensors are quite different. Therefore, at the training stage, we train our model without using the infrared and visible image pairs that are strictly registered. Aiming to verify that our method performs well in fusing image pairs without strict registration, we randomly translate the infrared images in the source image pairs with [-5,5], [-8,8]and [-10,10] pixels on the Roadscene dataset to get the misregistered infrared and visible image pairs, and then use the proposed method to fuse these

misregistered image pairs. The fusion results of these unregistered images are displayed in Figure 12. From these fusion results, the proposed method can preserve the target information from the source image. At the same time, the texture details from the source images are also fused into the fused image, which improves the quality of the fused image. The numbers in the red box of the fused images are still vivid, even under different unregistered degrees. The experimental results demonstrate the proposed method with strong robustness still has good performance in fusing images without registration.

4 Discussion

For the sake of avoiding the impact of changes in the agricultural working environment on the information perception for the intelligent agricultural system, we utilize infrared and visible image fusion to improve the image quality, so that the fused images can be used normally and even efficiently for various subsequent vision tasks in the intelligent agricultural system. Specifically, we propose a distributed fusion architecture for infrared and visible image fusion, termed RADFNet, which fuses images through three channels based on residual (RDCNN), edge attention, and multiscale channel attention. The proposed method can most retain the salient target information in the infrared image and the textural details information in the visible image. In addition, we introduce the multiscale channel attention module, which can extract coarse-to-fine features to preserve more information from source images to fused images. We also adopt an edge attention module that can compensate edge information for the fusedimage to make the fused image lose less

TABLE 2 The quantitative results on the RoadScene dataset with different a.

	a=1	a=10	a=50	a=100
EN	7.612254	7.604088	7.5805	7.58722
SF	0.088493	0.075895	0.076245	0.073816
SD	10.41727	10.50972	10.35863	10.38415
MI	2.670785	3.468535	3.501638	3.156152
VIF	0.698988	0.836942	0.832764	0.787895
AG	8.19272	7.033987	6.88297	6.939895

The red word represents the best, and the blue word represents the second best.
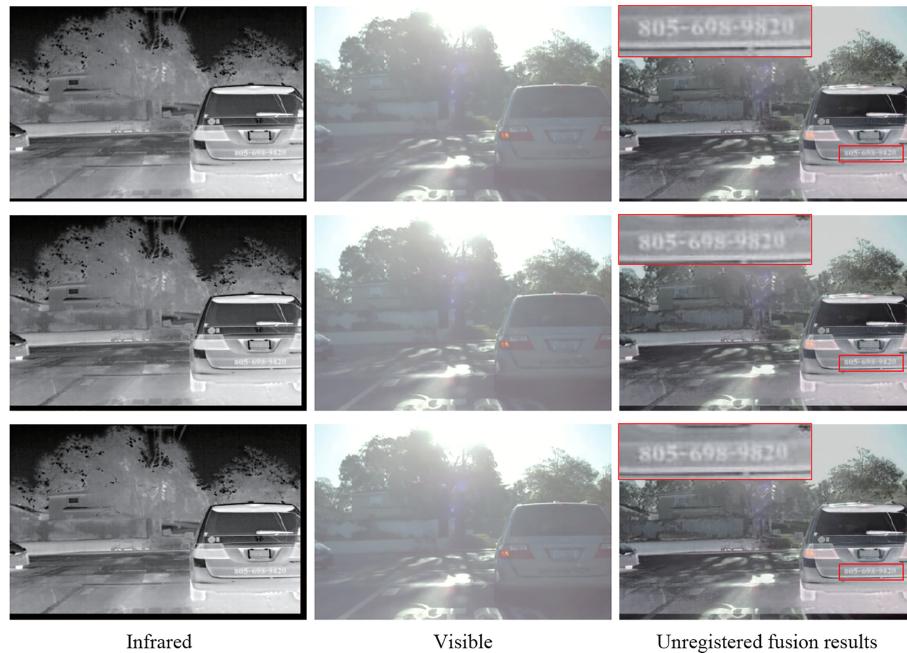
FIGURE 12

The results of fusing unregistered images with the proposed method on the Roadscene dataset. The infrared images are randomly translated, which causes the black edge in the images. The infrared images from top to bottom are translated with [-5,5],[-8,8],[-10,10] pixels.

edge information from source images. The comparative experiments are conducted on the Roadscene dataset, and the results demonstrate that the proposed method has superior performance in improving the fusion qualityand has achieved comparable results over the state-of-the-art image fusion algorithms in terms of visual effect and quantitative metrics. Finally, we send the unregistered image pairs into our network, and the results demonstrate that our method with strong robustness still performs well in fusing images without registration. The RADFNet performs well for infrared and visible image fusion due to the robust feature extraction ability of the network. The distributed fusion framework endows it with strong robustness, but the network parameters are still relatively large, which is not simple enough in the actual project deployment. In the future, it is necessary to improve the parameters of the network and the actual deployment of the model.

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

SF is responsible for the writing and theoretical design of the thesis, CW is responsible for the experimental test, and CL and MH are responsible for the verification of the experimental scheme, the analysis of the results and the project funding. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

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

Publisher's note

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

References

Aamir, M., Li, Z., Bazai, S., Wagan, R. A., Bhatti, U. A., Nizamani, M. M., et al. (2021). Spatiotemporal change of air-quality patterns in hubei province–a pre- to post-covid-19 analysis using path analysis and regression. *Atmosphere* 12. doi: 10.3390/atmos12101338atmos12101338

Basak, H., Kundu, R., Singh, P., Ijaz, M. F., Woźniak, M., and Sarkar, R. (2022). A union of deep learning and swarm-based optimization for 3d human action recognition. *Sci. Rep.* 12. doi: 10.1038/s41598-022-09293-8

Bavirisetti, D. P., Xiao, G., and Liu, G. (2017). "Multi-sensor image fusion based on fourth order partial differential equations," in *2017 20th International Conference on Information Fusion (Fusion)*. (Xi'an, China:IEEE) 1–9. doi: 10.23919/ICIF.2017.8009719

Bhatti, U. A., Ming-Quan, Z., Qing-Song, H., Ali, S., Hussain, A., Yuhuan, Y., et al. (2021). Advanced color edge detection using clifford algebra in satellite images. *IEEE Photonics J.* 13, 1–20. doi: 10.1109/JPHOT.2021.3059703

Bhatti, U. A., Nizamani, M. M., and Mengxing, H. (2022a). Climate change threatens pakistan's snow leopards. *Science* 377, 585–586. doi: 10.1126/science.add9065

Bhatti, U. A., Wu, G., Bazai, S. U., Ali Nawaz, S., Baryalai, M., Bhatti, M. A., et al. (2022b). A pre- to post-covid-19 change of air quality patterns in anhui province using path analysis and regression. *Polish J. Environ. Stud.* 31, 4029–4042. doi: 10.15244/pjoes/148065

Bhatti, U. A., Yu, Z., Chanussot, J., Zeeshan, Z., Yuan, L., Luo, W., et al. (2022c). Local similarity-based spatial–spectral fusion hyperspectral image classification with deep cnn and gabor filtering. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3090410

Bhatti, U. A., Zeeshan, Z., Nizamani, M. M., Bazai, S., Yu, Z., and Yuan, L. (2022d). Assessing the change of ambient air quality patterns in jiangsu province of china pre-to post-covid-19. *Chemosphere* 288, 132569. doi: 10.1016/j.chemosphere.2021.132569

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

Chen, Q., Ding, W., Huang, X., and Wang, H. (2022a). Generalized interval type ii fuzzy rough model based feature discretization for mixed pixels. *IEEE Trans. Fuzzy Syst.*, 1–15. doi: 10.1109/TFUZZ.2022.3190625

Chen, Q., Huang, M., and Wang, H. (2021). A feature discretization method for classification of high-resolution remote sensing images in coastal areas. *IEEE Trans. Geosci. Remote Sens.* 59, 8584–8598. doi: 10.1109/TGRS.2020.3016526

Chen, Q., Huang, M., Wang, H., and Xu, G. (2022b). A feature discretization method based on fuzzy rough sets for high-resolution remote sensing big data under linear spectral model. *IEEE Trans. Fuzzy Syst.* 30, 1328–1342. doi: 10.1109/TFUZZ.2021.3058020

Cui, G., Feng, H., Xu, Z., Li, Q., and Chen, Y. (2015). Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. *Optics Commun.* 341, 199–209. doi: 10.1016/j.optcom.2014.12.032

Da Cunha, A., Zhou, J., and Do, M. (2006). The nonsubsampled contourlet transform: Theory, design, and applications. *IEEE Trans. Image Process.* 15, 3089–3101. doi: 10.1109/TIP.2006.877507

Davis, J., and Sharma, V. (2007). Background-subtraction using contour-based fusion of thermal and visible imagery. *Comput. Vision Image Understanding* 106, 162–182. doi: 10.1016/j.cviu.2006.06.010

Eskicioglu, A., and Fisher, P. (1995). Image quality measures and their performance. *IEEE Trans. Commun.* 43, 2959–2965. doi: 10.1109/26.477498

Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.* 27, 1–10. doi: 10.1145/1360612.1360666

Gangapure, V. N., Nanda, S., and Chowdhury, A. S. (2018). Superpixel-based causal multisensor video fusion. *IEEE Trans. Circuits Syst. Video Technol.* 28, 1263–1272. doi: 10.1109/TCSVT.2017.2662743

Han, Y., Cai, Y., Cao, Y., and Xu, X. (2013). A new image fusion performance metric based on visual information fidelity. *Inf. Fusion* 14, –. doi: 10.1016/j.inffus.2011.08.002

He, K., Zhang, X., Ren, S., and Sun, J. (2016a). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Las Vegas, NV, USA:IEEE) 770–778. doi: 10.1109/CVPR.2016.90

He, K., Zhang, X., Ren, S., and Sun, J. (2016b). "Identity mappings in deep residual networks," in *Computer vision – ECCV 2016*. Eds. B. Leibe, J. Matas, N. Sebe and M. Welling (Cham: Springer International Publishing), 630–645.

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2020). Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023. doi: 10.1109/TPAMI.2019.2913372

Kingma, D. P., and Ba, J. (2015). . *Adam: A method for stochastic optimization. CoRR abs/1412.6980.*

Kong, W., Lei, Y., and Zhao, H. (2014). Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization. *Infrared Phys. Technol.* 67, 161–172. doi: 10.1016/j.infrared.2014.07.019

Li, J., Huo, H., Li, C., Wang, R., and Feng, Q. (2021b). Attentionfgan: Infrared and visible image fusion using attention-based generative adversarial networks. *IEEE Trans. Multimedia* 23, 1383–1396. doi: 10.1109/TMM.2020.2997127

Li, S., Kang, X., Fang, L., Hu, J., and Yin, H. (2017). Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* 33, 100–112. doi: 10.1016/j.inffus.2016.05.004

Li, H., and Wu, X.-J. (2018). Densefuse: A fusion approach to infrared and visible images. *IEEE Trans. Image Process.* 28, 2614–2623. doi: 10.1109/TIP.2018.2887342

Li, H., Wu, X.-J., and Durrani, T. (2020). Nestfuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. *IEEE Trans. Instrumentation Measurement* 69, 9645–9656. doi: 10.1109/TIM.2020.3005230

Li, H., Wu, X.-J., and Kittler, J. (2021a). Rfn-nest: An end-to-end residual fusion network for infrared and visible images. *Inf. Fusion* 73, 72–86. doi: 10.1016/j.inffus.2021.02.023

Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11, 674–693. doi: 10.1109/34.192463

Ma, J., Ma, Y., and Li, C. (2019a). Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* 45, 153–178. doi: 10.1016/j.inffus.2018.02.004

Ma, J., Xu, H., Jiang, J., Mei, X., and Zhang, X.-P. (2020). Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE Trans. Image Process.* 29, 4980–4995. doi: 10.1109/TIP.2020.2977573

Ma, J., Yu, W., Liang, P., Li, C., and Jiang, J. (2019b). Fusiongan: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* 48, 11–26. doi: 10.1016/j.inffus.2018.09.004

Ma, J., Zhang, H., Shao, Z., Liang, P., and Xu, H. (2021). Ganmcc: A generative adversarial network with multiclassification constraints for infrared and visible image fusion. *IEEE Trans. Instrumentation Measurement* 70, 1–14. doi: 10.1109/TIM.2020.3038013

Meng, F., Song, M., Guo, B., Shi, R., and Shan, D. (2017). Image fusion based on object region detection and non-subsampled contourlet transform. *Comput. Electr. Eng.* 62, 375–383. doi: 10.1016/j.compeleceng.2016.09.0191

Mitianoudis, N., Antonopoulos, S.-A., and Stathaki, T. (2013). "Region-based ica image fusion using textural information," in *2013 18th International Conference on Digital Signal Processing (DSP)*. (Fira, Greece:IEEE) 1–6. doi: 10.1109/ICDSP.2013.6622678

Mnih, V., Heess, N. M. O., Graves, A., and Kavukcuoglu, K. (2014). Recurrent models of visual attention. *ArXiv.* doi: 10.48550/arXiv.1406.6247

Naidu, D. V. (2011). Image fusion technique using multi-resolution singular value decomposition. *Defence Sci. J.* 61. doi: 10.14429/dsj.61.705

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on international conference on machine learning*, vol. 10. (Madison, WI, USA: Omnipress), 807–814. ICML.

Nencini, F., Garzelli, A., Baronti, S., and Alparone, L. (2007). Remote sensing image fusion using the curvelet transform. *Inf. Fusion* 8, 143–156. doi: 10.1016/j.inffus.2006.02.001

Prabhakar, K. R., Srikar, V. S., and Babu, R. V. (2017). "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *2017 IEEE International Conference on Computer Vision (ICCV)*. (Venice, Italy:IEEE) 4724–4732. doi: 10.1109/ICCV.2017.505

Qu, G., Zhang, D., and Yan, P. (2002). Information measure for performance of image fusion. *Electron. Lett.* 38, 313–315. doi: 10.1049/el:20020212

Ren, X., Meng, F., Hu, T., Liu, Z., and Wang, C. (2018). "Infrared-visible image fusion based on convolutional neural networks (cnn)," in *Intelligence science and big data engineering*. Eds. Y. Peng, K. Yu, J. Lu and X. Jiang (Cham: Springer International Publishing), 301–307.

Roberts, W., van Aardt, J., and Ahmed, F. (2008). Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *J. Appl. Remote Sens.* 2, 1–28. doi: 10.1117/1.2945910

Sun, S., Lin, H., Ma, J., and Li, X. (2017). Multi-sensor distributed fusion estimation with applications in networked systems: A review paper. *Inf. Fusion* 38, 122–134. doi: 10.1016/j.inffus.2017.03.006

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI*. (San Francisco, California, USA:AAAI Press).

Ulyanov, D., Vedaldi, A., and Lempitsky, V. S. (2016). Instance normalization: The missing ingredient for fast stylization. *ArXiv.* doi: 10.48550/arXiv.1607.08022

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *NIPS*. (Long Beach, California, USA:Curran Associates Inc. Conference Location).

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., et al. (2017). "Residual attention network for image classification," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Honolulu, HI, USA:IEEE) 6450–6458. doi: 10.1109/CVPR.2017.683

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

Wu, Y., Huang, M., Li, Y., Feng, S., and Wu, D. (2021). A Distributed Fusion Framework of Multispectral and Panchromatic Images Based on Residual Network. *Remote Sens.* 13 (13), 2556. doi: 10.3390/rs13132556

Xu, H., Ma, J., Jiang, J., Guo, X., and Ling, H. (2020). U2fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 502–518. doi: 10.1109/TPAMI.2020.3012548

Xu, H., Ma, J., Jiang, J., Guo, X., and Ling, H. (2022). U2fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 502–518. doi: 10.1109/TPAMI.2020.3012548

Yang, X., Zhang, W.-A., Yu, L., and Xing, K. (2016). Multi-rate distributed fusion estimation for sensor network-based target tracking. *IEEE Sensors J.* 16, 1233–1242. doi: 10.1109/JSEN.2015.2497464

Yan, G., and Woźniak, M. (2022). Accurate key frame extraction algorithm of video action for aerobics online teaching. *Mobile Networks Appl.* 27. doi: 10.1007/s11036-022-01939-1

Zhang, Q., Liu, Y., Blum, R. S., Han, J., and Tao, D. (2018). Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Information fusio*n. *Information Fusion* 40, 57–75. doi: 10.1016/j.inffus.2017.05.006

Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., and Zhang, L. (2020). Ifcnn: A general image fusion framework based on convolutional neural network. *Inf. Fusion* 54, 99–118. doi: 10.1016/j.inffus.2019.07.011

Zhang, H., and Ma, J. (2021). Sdnet: A versatile squeeze-and-decomposition network for real-time image fusion. *Int. J. Comput. Vision* 129. doi: 10.1007/s11263-021-01501-8

Zhang, X., Ma, Y., Fan, F., Zhang, Y., and Huang, J. (2017). Infrared and visible image fusion *via* saliency analysis and local edge-preserving multi-scale decomposition. . *J. Opt. Soc Am. A* 34, 1400–1410. doi: 10.1364/JOSAA.34.001400

Zhao, J.-X., Liu, J., Fan, D.-P., Cao, Y., Yang, J., and Cheng, M.-M. (2019a). "Egnet: Edge guidance network for salient object detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. (Seoul, Korea (South):IEEE) 8778–8787.

Zhao, W., Wang, D., and Lu, H. (2019b). Multi-focus image fusion with a natural enhancement *via* a joint multi-level deeply supervised convolutional neural network. *IEEE Trans. Circuits Syst. Video Technol.* 29, 1102–1115. doi: 10.1109/TCSVT.2018.2821177

Zhu, Z., Yin, H., Chai, Y., Li, Y., and Qi, G. (2018). A novel multi-modality image fusion method based on image decomposition and sparse representation. *Inf. Sci.* 432, 516–529. doi: 10.1016/j.ins.2017.09.010

frontiers | Frontiers in Plant Science

Evaluation of convolutional neural networks for herbicide susceptibility-based weed detection in turf

Xiaojun Jin[1,2], Teng Liu[2], Patrick E. McCullough[3], Yong Chen[1]* and Jialin Yu[2]*

[1]College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, Jiangsu, China, [2]Peking University Institute of Advanced Agricultural Sciences / Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang, Shandong, China, [3]Department of Crop and Soil Sciences, University of Georgia, Griffin, GA, United States

Deep learning methods for weed detection typically focus on distinguishing weed species, but a variety of weed species with comparable plant morphological characteristics may be found in turfgrass. Thus, it is difficult for deep learning models to detect and distinguish every weed species with high accuracy. Training convolutional neural networks for detecting weeds susceptible to herbicides can offer a new strategy for implementing site-specific weed detection in turf. DenseNet, EfficientNet-v2, and ResNet showed high F_1 scores (≥ 0.986) and MCC values (≥ 0.984) to detect and distinguish the sub-images containing dollarweed, goosegrass, old world diamond-flower, purple nutsedge, or Virginia buttonweed growing in bermudagrass turf. However, they failed to reliably detect crabgrass and tropical signalgrass due to the similarity in plant morphology. When training the convolutional neural networks for detecting and distinguishing the sub-images containing weeds susceptible to ACCase-inhibitors, weeds susceptible to ALS-inhibitors, or weeds susceptible to synthetic auxin herbicides, all neural networks evaluated in this study achieved excellent F_1 scores (≥ 0.995) and MCC values (≥ 0.994) in the validation and testing datasets. ResNet demonstrated the fastest inference rate and outperformed the other convolutional neural networks on detection efficiency, while the slow inference of EfficientNet-v2 may limit its potential applications. Grouping different weed species growing in turf according to their susceptibility to herbicides and detecting and distinguishing weeds by herbicide categories enables the implementation of herbicide susceptibility-based precision herbicide application. We conclude that the proposed method is an effective strategy for site-specific weed detection in turf, which can be employed in a smart sprayer to achieve precision herbicide spraying.

Introduction

Turfgrass is widely grown in urban landscapes, including institutional and residential lawns, parks, or athletic fields (Potter and Braman, 1991). The total turfgrass area in the United States is 163,812 km^2, which accounts for approximately 1.9% of the whole terrestrial land of the country (Milesi et al., 2005). Weed control is a challenging task for turfgrass management. Weeds compete with the turfgrass for sunlight, moisture, and soil nutrients, reducing turf aesthetics, surface quality, and functionality (Hamuda et al., 2016; Liu and Bruch, 2020). Weed management in turfgrass landscapes has relied heavily on broadcast herbicide application (McElroy and Martins, 2013), although weeds almost always present in non-uniform and patchy distributions (Dai et al., 2019; Yu et al., 2019a). Excessive application of synthetic herbicides could potentially pose a risk to human health and cause environmental pollution (Slaughter et al., 2008; Dai et al., 2019; Yu et al., 2019b; Mennan et al., 2020). Moreover, the application of synthetic herbicides represents a significant variable cost in turf weed management (Davis and Frisvold, 2017). These concerns have led to legal regulations regarding herbicide usage in several countries. For example, the European Union encourages spot-spraying to reduce the herbicide input (Busey, 2003; Marchand and Robin, 2019). Additionally, spot-spraying could effectively minimize the amount reaching off-target areas (Melland et al., 2016). In the United States, Environmental Protection Agency has proposed a series of measures, including prohibiting aerial applications for all atrazine labels to reduce their chance of runoff from the managed fields (Pimentel and Burgess, 2012; McCullough et al., 2015).

Site-specific weed management is a promising solution for sustainable weed control (Chen et al., 2022). Precision spraying a particular type or volume of herbicide onto susceptible weed species can significantly reduce herbicide input and weed control costs (Liakos et al., 2018). Site-specific weed management relies on the accurate identification and localization of weeds (Fennimore et al., 2016; Wang et al., 2019). Previous researchers explored various visual characteristics, such as color (Tang et al., 2016), morphological (Perez et al., 2000), hyper- or multi-spectral (Pantazi et al., 2016; Jiang et al., 2020), and textural features (Bakhshipour et al., 2017), for weed detection. However, crops and weeds may exhibit similar visual characteristics, thus detection and classification of weeds in crops are inherently challenging (Hasan et al., 2021). In turf, weed detection is challenging due to the presence of a variety of weed species growing with turfgrass.

In recent years, deep learning, a subfield of artificial intelligence, has demonstrated remarkable capability in speech recognition (Hinton et al., 2012; LeCun et al., 2015), natural language processing (Collobert and Weston, 2008; Collobert et al., 2011), and computer vision (Gu et al., 2018; Shi et al., 2020; Zhou et al., 2020). Deep learning technologies exhibit a tremendous ability to learn representations from raw data and extract complex features from images with a high accuracy level (Jordan and Mitchell, 2015; He et al., 2020; Yang et al., 2022a). Moreover, the improvements in graphics processing units (GPUs) have facilitated the use of deep convolutional neural networks (Bao et al., 2017; Bao et al., 2021; Ngo et al., 2021). Recent studies have investigated the feasibility of using deep learning in various agricultural domains, including plant disease detection (Martinelli et al., 2015; Saleem et al., 2019), crop yield prediction (Khaki and Wang, 2019; Van Klompenburg et al., 2020), plant phenotyping (Atefi et al., 2021; Zhang et al., 2022), and weed detection (Jin et al., 2021; Peng et al., 2022; Razfar et al., 2022). For example, Abbas et al. proposed a deep learning-based method for tomato disease detection. The trained neural network achieved a best 5-class classification accuracy of 99.51 (Abbas et al., 2021). Subeesh et al. compared four convolutional neural networks, including AlexNet, GoogLeNet, InceptionV3, and Xception for detecting various weeds growing in bell peppers (*Capsicum annum* L.) and found InceptionV3 achieved the highest accuracy (97.7%) (Subeesh et al., 2022). For image-based weed detection and discrimination, previous findings suggest that deep learning methods generally outperform other methods (Fennimore et al., 2016; Kamilaris and Prenafeta-Boldú, 2018).

Several studies have investigated the use of image classification or object detection neural networks for detecting and distinguishing various weed species in turfgrass (Yu et al., 2019a; Yu et al., 2019b; Yu et al., 2019c; Yu et al., 2020). Jin et al. demonstrated that VGGNet effectively detected and distinguished dallisgrass (*Paspalum dilataum* Poir.), purple nutsedge (*Cyperus rotundus* L.), and white clover (*Trifolium repens* L.) growing in bermudagrass [*Cynodon dactylon* (L.) Pers.] turf, while RegNet is well-performed in discriminating common dandelion (*Taraxacum officinale* Web.) (Jin et al., 2022). In another study, Yu et al. developed effective deep convolutional neural networks to detect weeds in turf. The authors reported that the image classification neural network VGGNet reliably classified broadleaf and grassy weeds growing in bermudagrass turf. In addition, the object detection neural network DetectNet achieved high overall accuracy at detecting cutleaf evening-primrose (*Oenothera laciniata* Hill) growing in bahiagrass (*Paspalum notatum* Flugge) (Yu et al., 2019b; Yu et al., 2019c).

Different weed species exhibit varying susceptibility to a particular herbicide category (McElroy and Martins, 2013; Yu et al., 2018). For example, acetolactate synthase (ALS)-inhibiting herbicides generally provide a narrow weed control spectrum (Yu and Boyd, 2018); ACCase-inhibiting herbicides are only effective for controlling grassy weeds (McElroy and Martins, 2013); nonselective herbicides, such as glyphosate and glufosinate, could nonselectively control all weeds (Johnson, 1977); and synthetic auxin herbicides, such as 2,4-D, dicamba, and MCPA, are only effective for controlling broadleaf weeds (McElroy and Martins, 2013). Therefore, precision spraying herbicides based on the susceptibility of different weed species to the herbicides can significantly reduce herbicide input and improve herbicide use efficiency. Although deep learning has been well-performed in weed detection and discrimination, previous studies have generally focused on distinguishing different weed species and did not establish a direct connection between weeds and herbicides. Moreover, a variety of weed species with comparable plant morphological characteristics may be found in turfgrass, thus it is difficult for the deep learning models to detect and distinguish every weed species with high accuracy. In the present research work, in addition to the detection and discrimination of individual weed species, different weed species growing in bermudagrass turf were grouped according to their susceptibility to herbicides, and weeds were detected and distinguished by herbicide categories. The proposed method would allow precision herbicide application based

on susceptibility and thereby effectively reduce herbicide input while achieving the same level of weed control as the broadcast herbicide application. The objectives of this paper were to (1) investigate the feasibility of utilizing deep learning for herbicide susceptibility-based weed detection in bermudagrass turf, and 2) evaluate and compare the performance of different convolutional neural networks for distinguishing individual weed species.

Materials and methods

Overview

The image classification convolutional neural networks, including DenseNet (Huang et al., 2017), EfficientNet (Tan and Le, 2019), and ResNet(He et al., 2016), were selected for evaluating the feasibility of using the convolutional neural networks for detecting and distinguishing individual weed species growing in bermudagrass turf or detecting and distinguishing weeds susceptible to herbicides. DenseNet is a convolutional neural network that computes dense and multi-scale features from the convolutional layers. For each layer, it obtains additional inputs from all preceding layers and passes on its feature maps to all subsequent layers. EfficientNet uses a set of fixed scaling coefficients to uniformly scales all dimensions of depth, width, and resolution in a principled way. The EfficientNet achieves state-of-the-art accuracy with 10× better efficiency by utilizing this novel scaling method. ResNet introduced the concept of residual learning. It employs an identity-based skip connection in each residual unit. ResNet eases the flow of information across units and thus can gain accuracy from very deep networks. In this study, these three convolutional neural networks were trained and evaluated with the ultimate goal of site-specific herbicide application.

Image acquisition

The training images of crabgrass (*D.igitaria ischaemum* L.), dollarweed (*Hydrocotyle* spp.), old world diamond-flower (*Hedyotis cormybosa* L.), and tropical signalgrass [*Urochloa distachya* (L.) T.Q. Nguyen] were acquired at several golf courses in Bradenton (27.49°N, 82.47°W), Riverview (27.86°N, 82.32°W), Sun City (27.71°N, 82.35° W), and Tampa (27.95°N, 82.45°W), Florida, while the testing images were acquired at several golf courses and institutional lawns in Lakeland, Florida (28.03°N, 81.94°W). The training images of goosegrass (*Eleusine indica* L.) and Virginia buttonweed (*Diodia virginiana* L.) growing in bermudagrass turf were acquired at the University of Georgia Griffin Campus in Griffin, Georgia, United States (33.26°N, 84.28°W), while the testing images were acquired at several golf courses in Peachtree City, Georgia, United States (33.39° N, 84.59°W). The training images of purple nutsedge were acquired at sod farms in Jiangning District, Nanjing, Jiangsu, China (31.95°N, 118.85°E), while the testing images were acquired at sod farms in Shuyang, Jiangsu, China (34.12°N, 118.79°E). The training and testing images of crabgrass, dollarweed, goosegrass, old world diamond-flower, tropical signalgrass, and Virginia buttonweed were captured multiple times from April to November 2018 using a digital camera (DSC-HX1, SONY®, Cyber-Shot Digital Still Camera, SONY

Corporation, Minato, Tokyo, Japan). The training and testing images of purple nutsedge were captured in spring 2021 using a digital camera (Panasonic® DMC-ZS110, Xiamen, Fujian, China). The original resolution of the training and testing images was 1,920 × 1,080 pixels. To enrich the diversity of the training dataset, images were captured under various illumination conditions, including partly cloudy, cloudy, or sunny days.

Training and testing

Images containing crabgrass, dollarweed, goosegrass, old world diamond-flower, purple nutsedge, tropical signalgrass, and Virginia buttonweed growing in bermudagrass turf were selected to constitute the training or testing datasets. Images containing a single weed species were selected for training and testing neural networks. All images were cropped into 40 equal-sized sub-images by a 5 rows × 8 columns division. Each sub-image was 240 × 216 pixels. Sub-images of crabgrass, goosegrass, and tropical signalgrass (Figure 1), purple nutsedge (Figure 2), dollarweed, old world diamond-flower, and Virginia buttonweed (Figure 3) at different growth stages and densities, and sub-images of bermudagrass at varying mowing heights and surface conditions (Figure 4) were utilized for training and testing the neural networks. Figure 5 outlines the sequence diagram of image processing and training and testing the convolutional neural networks for detecting and discriminating individual weed species or weeds susceptible to ACCase-inhibitor, ALS-inhibitor, synthetic auxin herbicides, or bermudagrass without weed infestation (no herbicide).

The convolutional neural networks for detecting and distinguishing weed species were trained utilizing a total of 21,000 true positive sub-images (3,000 sub-images for each weed species) containing crabgrass, dollarweed, goosegrass, old world diamond-flower, purple nutsedge, tropical signalgrass, or Virginia buttonweed growing in bermudagrass turf, while a total of 9,000 sub-images containing only bermudagrass were utilized as the true negative images. To establish the validation or testing dataset, a total of 3,500 sub-images (500 images for each weed species) containing crabgrass, dollarweed, goosegrass, old world diamond-flower, purple nutsedge, tropical signalgrass, or Virginia buttonweed growing in bermudagrass were utilized as the true positive images, while a total of 1,500 sub-images containing only bermudagrass were utilized as the true negative images.

The convolutional neural networks for detecting and distinguishing weeds susceptible to various herbicides were trained using a dataset containing four categories of sub-images: weed species susceptible to ACCase-inhibitors, weed species susceptible to ALS-inhibitors, weed species susceptible to synthetic auxin herbicides, and bermudagrass without weed infestation. To establish the training, validation, or testing dataset, the sub-images containing crabgrass, goosegrass, or tropical signalgrass, the sub-images containing purple nutsedge, the sub-images containing dollarweed, old world diamond-flower, or Virginia buttonweed, and the sub-images containing bermudagrass only were grouped and labeled as ACCase-inhibiting herbicides, ALS-inhibiting herbicides, synthetic auxin herbicides and no herbicide, respectively (Table 1).

The training and testing of the convolutional neural networks were performed in PyTorch (version 1.8.1) deep learning environment (Facebook, San Jose, California, United States) with an NVIDIA GeForce RTX 2080 Ti graphic processing unit (GPU).

FIGURE 1
The training and testing sub-images of crabgrass **(A)**, goosegrass **(B)**, and tropical signalgrass **(C)** at different growth stages and densities.

Transfer learning seeks to use previously acquired knowledge while addressing one problem and applying it to a different but similar problem (Lu et al., 2015). The convolutional neural networks were pre-trained with the ImageNet dataset to initialize the weights and bias through the transfer learning technology. To ensure fair comparisons among the evaluated deep learning models, default values of hyper-parameters for each neural network were adopted and used (Table 2).

A binary classification confusion matrix with four conditions, including the true positive (*tp*), false positive (*fp*), true negative (*tn*), and false negative (*fn*), was used to present the training and testing results of the convolutional neural networks. The performances of the convolutional neural networks were evaluated and compared against each other in terms of precision, recall, F_1 score, and Matthews Correlation Coefficient (MCC).

Precision is the ability of the neural networks to detect the susceptible weed species and was calculated using the *tp* and *fp* (Sokolova and Lapalme, 2009):

$$precision = \frac{tp}{tp+fp} \qquad (1)$$

Recall is the effectiveness of the neural networks to detect the susceptible weed species and was computed using the *tp* and *fn* (Sokolova and Lapalme, 2009):

$$recall = \frac{tp}{tp+fn} \qquad (2)$$

The F_1 score is a commonly used metric for measuring the overall performance of the neural networks, which was defined using the following equation (Sokolova and Lapalme, 2009):

$$F_1 = \frac{2 \times precision \times recall}{precision+recall} \qquad (3)$$

The MCC is the correlation between ground truth labels and predictions, which was determined using the following equation (Chicco and Jurman, 2020):

$$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp+fp) \times (tp+fn) \times (tn+fp) \times (tn+fn)}} \qquad (4)$$

Frames per second (FPS) measures the number of images, also known as frames processed by the neural networks each second. A higher FPS value indicates faster image processing. The FPS value was adopted as a quantitative metric to evaluate the computational efficiency of the neural networks.

Results

Detection and discrimination of weed species

When the convolutional neural networks were trained for detecting and distinguishing weed species growing in bermudagrass turf, DenseNet, EfficientNet-v2, and ResNet exhibited excellent

FIGURE 2
The training and testing sub-images of dollarweed **(A)**, old world diamond-flower **(B)**, and Virginia buttonweed **(C)** at different growth stages and densities.

performances and achieved high F_1 scores (≥0.995) and MCC values (≥0.994) in the validation datasets for detecting and distinguishing the sub-images containing dollarweed, goosegrass, purple nutsedge, and the sub-images containing bermudagrass only (Table 3). In general, a slight reduction in weed detection performance of all neural networks was observed in the testing datasets compared to the validation datasets.

For the detection of old world diamond-flower, the recall values of DenseNet in the validation and testing datasets were 0.994, while the precision values were 0.984 and 0.980, respectively, in predicting the correct weed species labels. For the detection of Virginia buttonweed, the precision values of DenseNet were 0.996 and 0.994, respectively, while the recall values were 0.984 and 0.978, respectively. Similar

trends were observed in the validation and testing datasets for EfficientNet-v2 and ResNet.

All three neural networks performed poorly at detecting and distinguishing crabgrass and tropical signalgrass growing in bermudagrass turf. Because of low precision and recall values, the F_1 scores and MCC values of DenseNet, EfficientNet-v2, and ResNet never exceeded 0.918, 0.919, and 0.918, respectively, in the validation and testing datasets. The low F_1 scores and MCC values indicate that the neural networks are more likely to mistakenly classify crabgrass as tropical signalgrass (or vice versa). This finding could likely attribute to the similarity in plant morphology between crabgrass and tropical signalgrass.

FIGURE 3
The training and testing sub-images of purple nutsedge at different growth stages and densities.

FIGURE 4
The training and testing sub-images of bermudagrass at different turfgrass management regimes, mowing heights, and surface conditions.

Detection and discrimination of weeds susceptible to herbicides

No obvious differences were observed among DenseNet, EfficientNet-v2, and ResNet for detecting and distinguishing weeds susceptible to ACCase-inhibitors, ALS-inhibitors, synthetic auxin herbicides, or bermudagrass without weed infestation (no herbicide) (Table 4).

DenseNet, EfficientNet-v2, and ResNet achieved high F_1 scores and MCC values (≥0.997) with high precision (≥0.996) and recall (≥0.997) in the validation datasets. All neural networks had slightly reduced precision and recall values in the testing datasets, but the F_1 scores and MCC values never fell below 0.994.

These results suggest that convolutional neural networks can reliably detect and distinguish weeds susceptible to particular herbicides. Furthermore, it can be inferred that training the neural networks based on the susceptibility of weed species to herbicides could probably minimize the morphological similarity issue and hence improve weed detection accuracy.

Inference time of the convolutional neural networks

In addition to the weed detection accuracy, the inference time of the convolutional neural networks is also critical for real-time precision herbicide application. The FPS values of DenseNet, EfficientNet-v2, and ResNet were calculated by averaging the inference time of images from the testing dataset. Since the original images were captured at a resolution of 1,920 × 1,080 pixels, the detection speed with the full images was measured by processing the sub-images (240 × 216 pixels) with a batch size value of 40 (for simultaneously processing 40 sub-images).
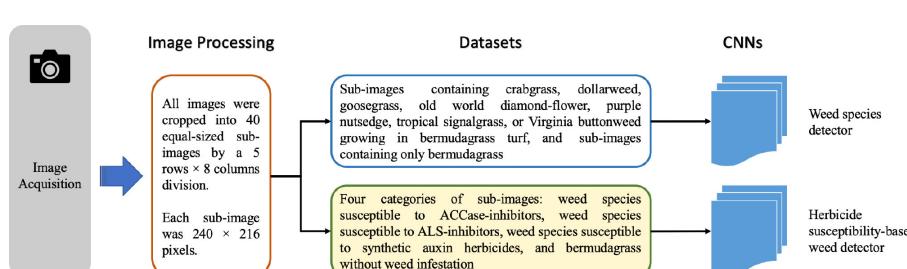

FIGURE 5
Flow diagram illustrates the sequence of image processing and training and testing the convolutional neural networks.

TABLE 1 The number of sub-images used to establish the training, validation, and testing datasets of the convolutional neural networks.

Dataset	ACCase-inhibiting herbicides			ALS-inhibiting herbicides	No herbicide	Synthetic auxin herbicides		
	Crabgrass	Goosegrass	Tropical signalgrass	Purple nut-sedge	Bermudagrass	Dollarweed	Old world diamond-flower	Virginia buttonweed
Training	3000	3000	3000	3000	9000	3000	3000	3000
Validation	500	500	500	500	1500	500	500	500
Testing	500	500	500	500	1500	500	500	500

The convolutional neural networks were trained to detect and discriminate weed species and the sub-images containing weeds susceptible to ACCase-inhibiting herbicides, ALS-inhibiting herbicides, synthetic auxin herbicides, or bermudagrass without weed infestation (no herbicide).

All convolutional neural networks, including DenseNet, EfficientNet-v2, and ResNet, had an excellent detection speed (≥77.94fps) when detecting and distinguishing the sub-images with a batch size value of 1 (Table 5). DenseNet, with 61.79 full images detected per second, was 31.59 slower than ResNet but noticeably faster than EfficientNet-v2 when setting the batch size value as 40. ResNet demonstrated the fastest inference rate and outperformed the other convolutional neural networks on detection efficiency. However, the slow detection of EfficientNet-v2 may limit its potential applications.

Discussion

Deep learning methods for weed detection typically focus on distinguishing weed species, but various weed species with comparable plant morphological features may be found in the turfgrass. Thus, it is difficult for neural networks to achieve high accuracy of detection and discrimination for every weed species. Distinguishing different categories of weed species growing in turf based on their susceptibility to herbicides reduces the complexity of weed detection. By training the neural networks according to the susceptibility of weed species to herbicides, we achieved an excellent performance in weed detection. Moreover, this strategy allows the use of specific herbicides for precision spraying susceptible weeds, thus saving more herbicides.

When training convolutional neural networks for detecting weeds susceptible to herbicides, weed vegetation was grouped and labeled into three categories: weeds susceptible to ACCase-inhibitors, weeds susceptible to ALS-inhibitors, and weeds susceptible to synthetic auxin herbicides. ACCase-inhibitors, such as diclofop-methyl, can be applied in bermudagrass turf for POST control of various grassy weeds, while sethoxydim (cyclohexanedione), another ACCase-inhibitor, is used for POST control of grassy weeds growing in centipedegrass [*Eremochloa ophiuroides* (Munro) Hack.] (Neal

et al., 1990; Tate et al., 2021). Synthetic auxin herbicides, such as 2,4-D and mecoprop, are POST herbicides that selectively control broadleaf weeds in bermudagrass turf (Grichar et al., 2008; Reed et al., 2013). ALS-inhibitors (e.g. halosulfuron, imazaquin, and trifloxysulfuron-sodium) can effectively control nutsedge weeds. However, it should be noted that certain ALS-inhibitors, such as halosulfuron and trifloxysulfuron-sodium, can also suppress or effectively control broadleaf weeds (McElroy and Martins, 2013). In this context, broadleaf and nutsedge weeds could be grouped into the same category when training the neural network for precision spraying the ALS-inhibitors that are effective for controlling both broadleaves and nutsedges growing in bermudagrass turf.

Deep learning neural networks, including image classification and object detection neural networks, can be developed and potentially integrated into the machine vision sub-system of a smart sprayer. Nevertheless, it should be noted that image classification neural networks alone do not localize weeds on the input images. Consequently, when utilizing image classification neural networks for weed detection, a smart sprayer likely generates a considerably larger spraying output area than the area covered by weeds. In the present work, localizing weeds with image classification neural networks could be realized by cropping the input image into multiple grid cells (sub-images) and identifying the grid cells containing weeds.

In the present study, original images (1,920 × 1,080 pixels) were divided into 40 grid cells (sub-images with a resolution of 240 × 216 pixels) for training and testing the image classification neural networks. Spraying areas can be localized by detecting if the grid cells contain weeds. When developing a precision spraying system, custom software can be programmed to generate grid cell maps on the input images and realize precision herbicide application by detecting if the grid cells contain weeds susceptible to the herbicides. To realize precision herbicide spraying, a binary (on/off) input command can be implemented *via* a nozzle control system to turn off the spray nozzles over the weed-free cells while the nozzles corresponding to the grid cells containing weeds need to be turned on.

TABLE 2 Hyperparameters used for training the convolutional neural networks.

Deep learning architecture	Optimizer	Base learning rate	Learning rate policy	Batch size	Training epochs
DenseNet	SGD	0.001	LambdaLR	16	30
EfficientNet-v2	SGD	0.01	LambdaLR	16	30
ResNet	Adam	0.0001	StepLR	16	30

SGD, stochastic gradient descent.

TABLE 3 Weed species detection and discrimination training results using various convolutional neural networks.

Deep learning architecture	Weed species	Validation dataset				Testing dataset			
		Precision	Recall	F_1 score	MCC	Precision	Recall	F_1 score	MCC
DenseNet	Bermudagrass	1.000	0.998	0.999	0.999	0.999	0.999	0.999	0.999
	Crabgrass	0.923	0.940	0.931	0.924	0.920	0.938	0.929	0.921
	Dollarweed	0.998	1.000	0.999	0.999	0.996	0.998	0.997	0.997
	Goosegrass	0.994	0.996	0.995	0.994	0.990	0.996	0.993	0.992
	Old world diamond-flower	0.984	0.994	0.989	0.988	0.980	0.994	0.987	0.986
	Purple nutsedge	0.994	0.998	0.996	0.996	0.996	0.994	0.995	0.994
	Tropical signalgrass	0.937	0.920	0.928	0.921	0.937	0.916	0.926	0.918
	Virginia buttonweed	0.996	0.984	0.990	0.989	0.994	0.978	0.986	0.984
EfficientNet-v2	Bermudagrass	1.000	1.000	1.000	1.000	1.000	0.999	0.999	1.000
	Crabgrass	0.924	0.942	0.933	0.925	0.920	0.938	0.929	0.921
	Dollarweed	1.000	1.000	1.000	1.000	0.998	0.998	0.998	0.998
	Goosegrass	0.998	0.996	0.997	0.997	0.992	0.996	0.994	0.993
	Old world diamond-flower	0.986	0.996	0.991	0.990	0.982	0.994	0.988	0.987
	Purple nutsedge	0.996	0.998	0.997	0.997	0.996	0.994	0.995	0.994
	Tropical signalgrass	0.941	0.922	0.931	0.924	0.937	0.918	0.927	0.919
	Virginia buttonweed	0.996	0.986	0.991	0.990	0.994	0.982	0.988	0.987
ResNet	Bermudagrass	0.999	0.999	0.999	0.998	0.999	0.999	0.999	0.999
	Crabgrass	0.922	0.942	0.932	0.924	0.918	0.938	0.928	0.920
	Dollarweed	1.000	0.998	0.999	0.999	0.998	0.998	0.998	0.998
	Goosegrass	0.998	0.996	0.997	0.997	0.990	0.996	0.993	0.992
	Old world diamond-flower	0.986	0.996	0.991	0.990	0.980	0.994	0.987	0.986
	Purple nutsedge	0.996	0.998	0.997	0.997	0.996	0.994	0.995	0.994
	Tropical signalgrass	0.941	0.918	0.929	0.921	0.937	0.916	0.926	0.918
	Virginia buttonweed	0.992	0.986	0.989	0.988	0.994	0.978	0.986	0.984

While the convolutional neural networks achieved high classification rates for detecting and distinguishing weeds susceptible to herbicides, it should be noted that when weeds susceptible to different herbicides are grown too close or occluded, the neural networks would not effectively distinguish weed categories based on their susceptibility to the herbicides because the grid cell contains multiple targets. Although such a case may result in missed detection, this is hardly an issue in field applications because the weed infestation zone has been detected, and one of the herbicides will be sprayed onto the susceptible weeds.

It was reported that the training image size could significantly affect the reliability of image classification neural networks for weed detection (Zhuang et al., 2021; Yang et al., 2022b). For example, Zhuang et al. observed increased classification accuracy (high recall values) with AlexNet and VGGNet when they were trained with images of 200 × 200 pixels than 300 × 300 or 400 × 400 pixels; however, increasing training image quantities diminished the differences in detection accuracy (Zhuang et al., 2021). In the present study, each sub-image (240 × 216 pixels) represented a physical size of 10 cm × 9 cm. When the convolutional neural networks are integrated into the machine vision sub-system of smart sprayers for precision herbicide application, the

nozzles should generate the same or slightly larger spraying outputs to cover the grid cells. An additional investigation is needed to investigate the implications of training image sizes and quantities on the performances of neural networks for weed detection in turf.

Conclusions

The present research demonstrated the reliability and effectiveness of using convolutional neural networks to detect and distinguish weeds growing in bermudagrass turf based on their susceptibility to herbicides. All convolutional neural networks, including DenseNet, EfficientNet-v2, and ResNet achieved excellent F_1 scores (≥ 0.995) and MCC values (≥ 0.994) in the validation and testing datasets to detect and distinguish weeds susceptible to ACCase-inhibitors, ALS-inhibitors, and synthetic auxin herbicides, or bermudagrass turf without weed infestation (no herbicide). In addition, DenseNet, EfficientNet-v2, and ResNet had an excellent detection speed (≥77.94fps) when detecting and distinguishing the sub-images with a resolution of 240 × 216 pixels. For detecting the original/full images (1,920 × 1,080 pixels), ResNet demonstrated the

TABLE 4   Training and testing results of various convolutional neural networks for detecting and discriminating the sub-images containing weeds susceptible to herbicides, or bermudagrass without weed infestation (no herbicide).

Deep learning architecture	Herbicides	Validation dataset				Testing dataset			
		Precision	Recall	F$_1$ score	MCC	Precision	Recall	F$_1$ score	MCC
DenseNet	ACCase-inhibiting herbicides	0.999	0.999	0.999	0.998	0.997	0.998	0.997	0.997
	ALS-inhibiting herbicides	0.996	0.998	0.997	0.997	0.996	0.994	0.995	0.994
	No herbicide	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
	Synthetic auxin herbicides	0.999	0.999	0.999	0.998	0.999	0.999	0.999	0.999
EfficientNet-v2	ACCase-inhibiting herbicides	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.997
	ALS-inhibiting herbicides	0.996	0.998	0.997	0.997	0.996	0.994	0.995	0.994
	No herbicide	0.999	0.999	0.999	0.999	1.000	0.999	0.999	1.000
	Synthetic auxin herbicides	0.999	0.999	0.999	0.999	0.999	1.000	0.999	0.999
ResNet	ACCase-inhibiting herbicides	0.999	0.998	0.998	0.998	0.998	0.995	0.996	0.995
	ALS-inhibiting herbicides	0.996	0.998	0.997	0.997	0.996	0.994	0.995	0.994
	No herbicide	0.998	1.000	0.999	0.999	0.997	0.999	0.998	0.997
	Synthetic auxin herbicides	0.999	0.997	0.998	0.998	0.997	0.997	0.997	0.996

TABLE 5   The inference time of the convolutional neural networks evaluated in the study.

Deep learning architecture	Image type	Resolution	Batch size	FPS
DenseNet	Sub-image	240 × 216	1	103.75
			40	61.79
EfficientNet-v2	Sub-image	240 × 216	1	77.94
			40	38.77
ResNet	Sub-image	240 × 216	1	276.08
			40	93.38

FPS, frames per second.

fastest inference rate and outperformed the other convolutional neural networks on detection efficiency (93.38fps). Effective detection and discrimination of weeds susceptible to herbicides enable the smart sprayer to spray particular herbicides to control susceptible weeds, thereby significantly reducing herbicide input. Based on the high-level performance, we conclude that the proposed method is highly suitable for integrating into the machine vision sub-system of smart sprayers for the precision control of weeds while growing in turf.

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

XJ conceived the research ideas and designed the experiments under the guidance of YC and JY. TL, PM, and JY collected the data and conducted the data analysis. XJ drafted the manuscript. PM, YC, and JY edited and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

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

Publisher's note

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

References

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, 106279. doi: 10.1016/j.compag.2021.106279

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. doi: 10.3389/fpls.2021.611940

Bakhshipour, A., Jafari, A., Nassiri, S. M., and Zare, D. (2017). Weed segmentation using texture features extracted from wavelet sub-images. *Biosyst. Eng.* 157, 1–12. doi: 10.1016/j.biosystemseng.2017.02.002

Bao, W., Yang, B., and Chen, B. (2021). 2-hydr_ensemble: lysine 2-hydroxyisobutyrylation identification with ensemble method. *Chemom. Intell. Lab. Syst.* 215, 104351. doi: 10.1016/j.chemolab.2021.104351

Bao, W., Yuan, C.-A., Zhang, Y., Han, K., Nandi, A. K., Honig, B., et al. (2017). Mutli-features prediction of protein translational modification sites. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15 (5), 1453–1460. doi: 10.1109/TCBB.2017.2752703

Busey, P. (2003). Cultural management of weeds in turfgrass: a review. *Crop Sci.* 43 (6), 1899–1911. doi: 10.2135/cropsci2003.1899

Chen, D., Lu, Y., Li, Z., and Young, S. (2022). Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. *Comput. Electron. Agric.* 198, 107091. doi: 10.1016/j.compag.2022.107091

Chicco, D., and Jurman, G. (2020). The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21 (1), 1–13. doi: 10.1186/s12864-019-6413-7

Collobert, R., and Weston, J. (2008). "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning* (Helsinki, Finland:ACM). 160–167.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* 12 (ARTICLE), 2493–2537.

Dai, X., Xu, Y., Zheng, J., and Song, H. (2019). Analysis of the variability of pesticide concentration downstream of inline mixers for direct nozzle injection systems. *Biosyst. Eng.* 180, 59–69. doi: 10.1016/j.biosystemseng.2019.01.012

Davis, A. S., and Frisvold, G. B. (2017). Are herbicides a once in a century method of weed control? *Pest Manage. science.* 73 (11), 2209–2220. doi: 10.1002/ps.4643

Fennimore, S. A., Slaughter, D. C., Siemens, M. C., Leon, R. G., and Saber, M. N. (2016). Technology for automation of weed control in specialty crops. *Weed Technol.* 30 (4), 823–837. doi: 10.1614/WT-D-16-00070.1

Grichar, W. J., Baumann, P. A., Baughman, T. A., and Nerada, J. D. (2008). Weed control and bermudagrass tolerance to imazapic plus 2, 4-d. *Weed Technol.* 22 (1), 97–100. doi: 10.1614/WT-07-097.1

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recogn.* 77, 354–377. doi: 10.1016/j.patcog.2017.10.013

Hamuda, E., Glavin, M., and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* 125, 184–199. doi: 10.1016/j.compag.2016.04.024

Hasan, A. M., Sohel, F., Diepeveen, D., Laga, H., and Jones, M. G. (2021). A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* 184, 106067. doi: 10.1016/j.compag.2021.106067

He, T., Liu, Y., Yu, Y., Zhao, Q., and Hu, Z. (2020). Application of deep convolutional neural network on feature extraction and detection of wood defects. *Measurement* 152, 107357. doi: 10.1016/j.measurement.2019.107357

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, USA:IEEE). 770–778.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Magazine.* 29 (6), 82–97. doi: 10.1109/MSP.2012.2205597

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, USA:IEEE). 4700–4708.

Jiang, H., Jiang, X., Ru, Y., Wang, J., Xu, L., and Zhou, H. (2020). Application of hyperspectral imaging for detecting and visualizing leaf lard adulteration in minced pork. *Infrared Phys. Technology.* 110, 103467. doi: 10.1016/j.infrared.2020.103467

Jin, X., Bagavathiannan, M., McCullough, P. E., Chen, Y., and Yu, J. (2022). A deep learning-based method for classification, detection, and localization of weeds in turfgrass. *Pest Manage. Sci* 78(11), 4809-4821. doi: 10.1002/ps.7102

Jin, X., Che, J., and Chen, Y. (2021). Weed identification using deep learning and image processing in vegetable plantation. *IEEE Access.* 9, 10940–10950. doi: 10.1109/ACCESS.2021.3050296

Johnson, B. (1977). Winter annual weed control in dormant bermudagrass turf. *Weed Sci.* 25 (2), 145–150. doi: 10.1017/S0043174500033142

Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science* 349 (6245), 255–260. doi: 10.1126/science.aaa8415

Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016

Khaki, S., and Wang, L. (2019). Crop yield prediction using deep neural networks. *Front. Plant science.* 10, 621. doi: 10.3389/fpls.2019.00621

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553), 436–444. doi: 10.1038/nature14539

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors* 18 (8), 2674. doi: 10.3390/s18082674

Liu, B., and Bruch, R. (2020). Weed detection for selective spraying: a review. *Curr. Robotics Rep.* 1 (1), 19–26. doi: 10.1007/s43154-020-00001-w

Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., and Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems.* 80, 14–23. doi: 10.1016/j.knosys.2015.01.010

Marchand, P. A., and Robin, D. (2019). Evolution of directive (EC) no 128/2009 of the European parliament and of the council establishing a framework for community action to achieve the sustainable use of pesticides. *J. Regul. Sci.*, 7, 1–7. doi: 10.21423/jrs-v07marchand

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., et al. (2015). Advanced methods of plant disease detection. *A review. Agron. Sustain. Dev* 35 (1), 1–25. doi: 10.1007/s13593-014-0246-1

McCullough, P. E., Yu, J., Shilling, D. G., Czarnota, M. A., and Johnston, C. R. (2015). Biochemical effects of imazapic on bermudagrass growth regulation, broomsedge (Andropogon virginicus) control, and MSMA antagonism. *Weed Sci.* 63 (3), 596–603. doi: 10.1614/WS-D-14-00183.1

McElroy, J., and Martins, D. (2013). Use of herbicides on turfgrass. *Planta daninha.* 31, 455–467. doi: 10.1590/S0100-83582013000200024

Melland, A. R., Silburn, D. M., McHugh, A. D., Fillols, E., Rojas-Ponce, S., Baillie, C., et al. (2016). Spot spraying reduces herbicide concentrations in runoff. *J. Agric. Food Chem.* 64 (20), 4009–4020. doi: 10.1021/acs.jafc.5b03688

Mennan, H., Jabran, K., Zandstra, B. H., and Pala, F. (2020). Non-chemical weed management in vegetables by using cover crops: A review. *Agronomy* 10 (2), 257. doi: 10.3390/agronomy10020257

Milesi, C., Elvidge, C., Dietz, J., Tuttle, B., Nemani, R., and Running, S. (2005). A strategy for mapping and modeling the ecological effects of US lawns. *J. Turfgrass Manage.* 1 (1), 83–97.

Neal, J. C., Bhowmik, P. C., and Senesac, A. F. (1990). Factors influencing fenoxaprop efficacy in cool-season turfgrass. *Weed Technol.* 4 (2), 272–278. doi: 10.1017/S0890037X00025380

Ngo, T. D., Bui, T. T., Pham, T. M., Thai, H. T., Nguyen, G. L., and Nguyen, T. N. (2021). Image deconvolution for optical small satellite with deep learning and real-time GPU acceleration. *J. Real-Time Image Processing.* 18 (5), 1697–1710. doi: 10.1007/s11554-021-01113-y

Pantazi, X.-E., Moshou, D., and Bravo, C. (2016). Active learning system for weed species recognition based on hyperspectral sensing. *Biosyst. Eng.* 146, 193–202. doi: 10.1016/j.biosystemseng.2016.01.014

Peng, H., Li, Z., Zhou, Z., and Shao, Y. (2022). Weed detection in paddy field using an improved RetinaNet network. *Comput. Electron. Agric.* 199, 107179. doi: 10.1016/j.compag.2022.107179

Perez, A., Lopez, F., Benlloch, J., and Christensen, S. (2000). Colour and shape analysis techniques for weed detection in cereal fields. *Comput. Electron. Agric.* 25 (3), 197–212. doi: 10.1016/S0168-1699(99)00068-X

Pimentel, D., and Burgess, M. (2012). Small amounts of pesticides reaching target insects. *Springer*, 14, 1–2. doi: 10.1007/s10668-011-9325-5

Potter, D. A., and Braman, S. K. (1991). Ecology and management of turfgrass insects. *Annu. Rev. Entomol.* 36 (1), 383–406. doi: 10.1146/annurev.en.36.010191.002123

Razfar, N., True, J., Bassiouny, R., Venkatesh, V., and Kashef, R. (2022). Weed detection in soybean crops using custom lightweight deep learning models. *J. Agric. Food Res.* 8, 100308. doi: 10.1016/j.jafr.2022.100308

Reed, T. V., Yu, J., and McCullough, P. E. (2013). Aminocyclopyrachlor efficacy for controlling Virginia buttonweed (Diodia virginiana) and smooth crabgrass (Digitaria ischaemum) in tall fescue. *Weed Technol.* 27 (3), 488–491. doi: 10.1614/WT-D-12-00159.1

Saleem, M. H., Potgieter, J., and Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants* 8 (11), 468. doi: 10.3390/plants8110468

Shi, J., Li, Z., Zhu, T., Wang, D., and Ni, C. (2020). Defect detection of industry wood veneer based on NAS and multi-channel mask r-CNN. *Sensors* 20 (16), 4398. doi: 10.3390/s20164398

Slaughter, D. C., Giles, D. K., and Downey, D. (2008). Autonomous robotic weed control systems: A review. *Comput. Electron. Agric.* 61 (1), 63–78. doi: 10.1016/j.compag.2007.05.008

Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inf. Process. management.* 45 (4), 427–437. doi: 10.1016/j.ipm.2009.03.002

Subeesh, A., Bhole, S., Singh, K., Chandel, N., Rajwade, Y., Rao, K., et al. (2022). Deep convolutional neural network models for weed detection in polyhouse grown bell peppers. *Artif. Intell. Agriculture.* 6, 47–54. doi: 10.1016/j.aiia.2022.01.002

Tang, J.-L., Chen, X.-Q., Miao, R.-H., and Wang, D. (2016). Weed detection using image processing under different illumination for site-specific areas spraying. *Comput. Electron. Agric.* 122, 103–111. doi: 10.1016/j.compag.2015.12.016

Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning. PMLR* (Long Beach, USA:PMLR). 6105–6114.

Tate, T. M., McCullough, P. E., Harrison, M. L., Chen, Z., and Raymer, P. L. (2021). Characterization of mutations conferring inherent resistance to acetyl coenzyme a carboxylase-inhibiting herbicides in turfgrass and grassy weeds. *Crop Sci.* 61 (5), 3164–3178. doi: 10.1002/csc2.20511

Van Klompenburg, T., Kassahun, A., and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* 177, 105709. doi: 10.1016/j.compag.2020.105709

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

Yang, J., Bagavathiannan, M., Wang, Y., Chen, Y., and Yu, J. (2022b). A comparative evaluation of convolutional neural networks, training image sizes, and deep learning optimizers for weed detection in alfalfa. *Weed Technol.* 36 (4), 512–522. doi: 10.1017/wet.2022.46

Yang, B., Bao, W., and Wang, J. (2022a). Active disease-related compound identification based on capsule network. *Brief. Bioinform.* 23 (1), bbab462. doi: 10.1093/bib/bbab462

Yu, J., and Boyd, N. S. (2018). Weed control and tolerance of sulfonylurea herbicides in caladium. *Weed Technol.* 32 (4), 424–430. doi: 10.1017/wet.2018.30

Yu, J., McCullough, P. E., and Czarnota, M. A. (2018). Annual bluegrass (Poa annua) biotypes exhibit differential levels of susceptibility and biochemical responses to protoporphyrinogen oxidase inhibitors. *Weed Sci.* 66 (5), 574–580. doi: 10.1017/wsc.2018.30

Yu, J., Schumann, A. W., Cao, Z., Sharpe, S. M., and Boyd, N. S. (2019a). Weed detection in perennial ryegrass with deep learning convolutional neural network. *Front. Plant science.* 10. doi: 10.3389/fpls.2019.01422

Yu, J., Schumann, A. W., Sharpe, S. M., Li, X., and Boyd, N. S. (2020). Detection of grassy weeds with deep convolutional neural networks. *Weed Sci.* 68 (5), 545–552. doi: 10.1017/wsc.2020.46

Yu, J., Sharpe, S. M., Schumann, A. W., and Boyd, N. S. (2019b). Deep learning for image-based weed detection in turfgrass. *Eur. J. Agron.* 104, 78–84. doi: 10.1016/j.eja.2019.01.004

Yu, J., Sharpe, S. M., Schumann, A. W., and Boyd, N. S. (2019c). Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. *Pest Manage. Science.* 75 (8), 2211–2218. doi: 10.1002/ps.5349

Zhang, H., Ge, Y., Xie, X., Atefi, A., Wijewardane, N. K., and Thapa, S. (2022). High throughput analysis of leaf chlorophyll content in sorghum using RGB, hyperspectral, and fluorescence imaging and sensor fusion. *Plant Methods* 18 (1), 1–17. doi: 10.1186/s13007-022-00892-0

Zhou, H., Zhuang, Z., Liu, Y., Liu, Y., and Zhang, X. (2020). Defect classification of green plums based on deep learning. *Sensors* 20 (23), 6993. doi: 10.3390/s20236993

Zhuang, J., Li, X., Bagavathiannan, M., Jin, X., Yang, J., Meng, W., et al. (2021). Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. *Pest Manage. Sci* 78 (5), 1861-1869. doi: 10.1002/ps.6656

Check for updates

BAF-Net: Bidirectional attention fusion network *via* CNN and transformers for the pepper leaf segmentation

Jiangxiong Fang[1], Houtao Jiang[2], Shiqing Zhang[1], Lin Sun[2], Xudong Hu[2,3], Jun Liu[4], Meng Gong[2], Huaxiang Liu[1*] and Youyao Fu[1*]

[1]Institute of Intelligent Information Processing, Taizhou University, Taizhou, Zhejiang, China, [2]Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology, East China University of Technology, Nanchang, China, [3]Engineering Research Center of Development and Management for Low to Ultra-Low Permeability Oil & Gas Reservoirs in West China, Xi'an Shiyou University, Xi'an, China, [4]College of Mechanical Engineering, Quzhou University, Quzhou, Zhejiang, China

The segmentation of pepper leaves from pepper images is of great significance for the accurate control of pepper leaf diseases. To address the issue, we propose a bidirectional attention fusion network combing the convolution neural network (CNN) and Swin Transformer, called BAF-Net, to segment the pepper leaf image. Specially, BAF-Net first uses a multi-scale fusion feature (MSFF) branch to extract the long-range dependencies by constructing the cascaded Swin Transformer-based and CNN-based block, which is based on the U-shape architecture. Then, it uses a full-scale feature fusion (FSFF) branch to enhance the boundary information and attain the detailed information. Finally, an adaptive bidirectional attention module is designed to bridge the relation of the MSFF and FSFF features. The results on four pepper leaf datasets demonstrated that our model obtains F1 scores of 96.75%, 91.10%, 97.34% and 94.42%, and IoU of 95.68%, 86.76%, 96.12% and 91.44%, respectively. Compared to the state-of-the-art models, the proposed model achieves better segmentation performance. The code will be available at the website: https://github.com/fangchj2002/BAF-Net.

KEYWORDS

convolution neural network, leaf segmentation, attention mechanism, multi-scale network, Swin Transformer

1 Introduction

Pepper is a common crop in China and has become an important vegetable and condiment in our daily life. However, pepper is a sensitive plant and pepper crops are highly exposed to diseases, which easily cause the frontal disease of the pepper leaves. The plant leaves can reflect plant growth, and pepper leaf diseases directly leads to the decline of

pepper yield and quality. The visual characteristics of pepper leaf diseases is very similar, so it is not easy to distinguish them. With the advance of imaging technology, computer vision technologies have been widely used in plant leaf extraction to guide the agricultural expert to analyze the crop growth. By using image processing technology to analyze two-dimensional leaf image features, the plant growth stages could be dissected (Slaughter et al., 2008; Koirala et al., 2019), and monitor the plant diseases (Singh, 2019; Tian et al., 2019) by the analysis of the image various plant organs. Therefore, the accurate segmentation of pepper leaves from pepper images is of great significance for controlling pepper leaf diseases. However, it is challenging to design a general model for automatic segmentation of pepper leaves since the pepper leaves and some crops have similar phenotypic features (Hasan et al., 2021).

Broadly speaking, the existing literature for the plant leaf segmentation can be classified into two categories as shown in Figure 1: conventional and deep learning-based methods. For the conventional methods, a statistical method with graph-based models (Kumar and Domnic, 2019) was proposed to segment the plant image and leaf counting, where the image enhancement techniques and the transformation from RGB to HSV were used to improve the quality of the input image. To avoid the problem of leaf over-segmentation, green channel information (Wang et al., 2018) was used to remove the background information, and the Sobel operator was improved to segment cucumber leaves. To detect the occluded plant leaves, leaf shape (Xia et al., 2013) was fused into the energy function to segment the leaf images. To deal with the complex background and the strong illumination, Larese et al. (2014) proposed a leaf vein analysis method for leguminous leaf segmentation and classification. The automatic segmentation method for plant leaf images under complex background was proposed to obtain the segmentation results. Scharr et al. (2016) uses the supervised classification with a neural network along with color and watershed transform for plant leaf segmentation and counting. Kuznichov et al. (2019) proposed a schema to augment the training dataset and remain the geometrical structure of the plant leaf by constructing a generation synthetic data. To segment multiple leaves at the same time and deal with the leaf over-segmentation, a deep extraction method for plant leaf (Amean et al., 2021) was proposed by incorporating multiple features, such as color, shape, and depth information. Lin et al. (Lin et al., 2023) proposed a self-supervised blade segmentation framework consisting of a self-supervised semantic segmentation model, a color-based blade segmentation algorithm, and a self-supervised color correction model. A self-supervised semantic segmentation model (Lin et al., 2023) was proposed to deal with the complex lighting conditions. The model was comprised of the features extracted from the CNN-based network and the fully connected Conditional Random Fields (CRFs), thus significantly reducing the impact of complex backgrounds and variations within the leaf and non-leaf regions.

In recent years, the deep learning-based method has outperformed the conventional segmentation methods and shows great potential in processing plant phenotypic tasks (Bhagat et al., 2021; Chandra et al., 2020). The SegNet-based model (Aich and Stavness, 2017) with the encoder-decoder architecture was used to segment plant leaves and leaf counting. Three RGB images and the segmentation mask of leaf counting were used as four input channels to build a regression model. Thus, the SegNet-based model can solve the problem of leaf counting (Ubbens and Stavness, 2017). To segment multiple objects, the instance segmentation model (Romera-Paredes and Torr, 2016) was proposed based on an end-to-end recurrent neural network (RNN). The model designed a spatial attention module to extract small patches, and then uses a convolutional long short-term memory (LSTM) network to build the relation of these patches. By doing so, the model can finish plant leaf segmentation and leaf counting. To solve the target occlusion problem, Ren et al. (Ren and Zemel, 2017) used an RNN-based architecture to generate continuous regions of interest and designed a human-like counting process based on the attention mechanism, thus making it a more accurate segmentation for each object in turn. Lin et al. (Lin et al., 2019) proposed a self-supervised CNN-based framework for leaf segmentation. The model first used self-contained information to classify each pixel, and then the segmentation algorithm for the color leaf images was used to identify the leaf region. Finally, a self-supervised color-based correction model was proposed to segment the complex images taken under complex lighting conditions. As shown in Table 1, we summarize the work related to plant leaf segmentation.

It is well-known that U-Net (Ronneberger et al., 2015) is one of the most efficient models and widely used for specific object extraction in image segmentation. U-Net and its variants (Shen et al., 2017) have achieved competitive performance in many computer vision tasks, such as ResU-Net (Zhang et al., 2018), U-Net++ (Zhou et al., 2019), DenseNet (Huang et al., 2017), 3D U-Net (Li et al., 2020), V-Net (Milletari et al., 2016). Bhagat et al. (2022) proposed a modified U-Net architecture for plant leaf segmentation, where an EffcientNet-B4 module was used as an encoder to extract the image feature. Meanwhile, a redesigned skip connection and the residual modules of the decoder were used to reduce computational cost. However, these methods usually ignored the global context information. To be exact, these models could not extract the long-range correlation between pixels, especially for the pixels surrounding the boundary of the objects. The effective method for obtaining the precise location and boundary of the segmentation object was to extract the global context information of the feature map and the long-range correlation between pixels. Transformer has been proved to be an efficient self-attention mechanism to establish long-term dependencies in the field of natural language processing (NLP). More recently, it was introduced into the visual classification tasks. Ramachandran et al. (Ramachandran et al., 2019) explored a novel ResNet-based model by replacing all spatial convolutional layers with the self-attention layers. However, the local self-attention might still lose part of the global structural information. In order to obtain global information of visual images, Vision Transformer (ViT) (Dosovitskiy et al., 2020) inspired by Transformer was proposed to solve the natural image recognition task. ViT first divided the image into several non-overlapping patches, and then used Transformer with the self-attention mechanism to calculate the global information between each token to obtain the global context information. To further reduce the sequence length and computational complexity, Swin

TABLE 1 The related works in plant leaf image.

Categories	Author	Method
Conventional method	Kumar and Domnic, 2019	A statistical method with graph-based models
	Wang et al., 2018	The Sobel-based model with green channel information
	Xia et al., 2013	The modified active shape models for plant leaf detection
	Larese et al., 2014	Automatic classification modle for legumes image
	Kuznichov et al., 2019	Augment dataset and the geometrical structure
	Amean et al., 2021	Self-supervised blade segmentation framework
	Lin et al., 2023	Self-supervised semantic segmentation model for complex lighting conditions
Deep learning-based method	Aich and Stavness, 2017	The SegNet-based model for leaves and leaf counting
	Ubbens and Stavness, 2017	A deep learning platform for complex plant phenotyping
	Romera-Paredes and Torr, 2016	Recurrent instance segmentation
	Ren and Zemel, 2017	End-to-end instance segmentation with recurrent attention
	Lin et al., 2019	A self-supervised CNN-based framework for leaf segmentation

Transformer (Liu et al., 2021) used a shifted window to calculate the local self-attention. By establishing a shifted window, two adjacent windows could interact with each other, and cross connections were established between the widows of the upper and lower layers, which improved the effect of global context.

To address these problems, we built a pepper leaf dataset focused on the disease detection segmentation, and propose a bidirectional attention fusion network, named BAF-Net, to obtain the for pepper leaf segmentation. BAF-Net is comprised of three parts: multi-scale fusion feature (MSFF) branch, full-scale feature fusion (FSFF) branch, and bidirectional attention feature fusion (BAF) modules. The backbone of the MSFF branch is a U-shaped network architecture. By incorporating the Swin-Transformer block and the CNN-based module, a cascaded hybrid module (Swin-Trans-Conv) is constructed, to obtain multi-scale fusion features. In the FSFF branch, we first fuse the features of the five-layer encoder from the MSFF branch. Then, the generated features pass through several convolution blocks to obtain the full-resolution feature. The BAF module adaptively fuses the output features of the MSFF and FSFF branches, generating two corresponding features for each branch. In short, the main contributions of our work are as follows:

(1) By incorporating the Swin Transformer and CNN-based modules, we build a cascaded Swin-Trans-Conv block to replace each convolutional layer of U-Net. The Swin Transformer-based module can extract the long-range dependencies while the CNN-based module is used to obtain the local image information.

(2) An FSFF branch is designed to extract detailed information and the boundaries. By incorporating the multi-scale features which are from the outputs of the encoder in the MSFF branch, the boundary information is retained. Meanwhile, the multi-layer full-scale convolution block can extract detailed information.

(3) We propose a BAF module to adaptive share the multi-scale and full-scale features, which can adaptively compute the features of two corresponding branches according to the output features of the MSFF and FSFF branches.

(4) By verifying on four dataset of pepper leaf images, the results show that our model is superior to the state-of-the-art models in terms of the evaluation indices such as IoU and F1score.

The rest of this paper is arranged as follows. Section 2 first reviews the materials including the dataset and its labeling process. Then, the proposed model including the overall architecture, the formulation of the MSFF and FSFF branches, the BAF module, and the loss function are discussed. Finally we introduce the evaluation indices. Section 3 demonstrates the experimental results and discussion. The conclusions are summarized in Section 4.

2 Materials and methods

2.1 Dataset

In our experiments, the images of pepper leaves were taken from the farm of Nanchang Academy of Agricultural Sciences in Jiangxi Province, China. We took photos for multi-view in the real natural environment from the morning to the afternoon on August 12 and 13, 2021.Pepper leaves were seriously affected by a variety of diseases during growth. Two common diseases of pepper leaf destroyed the normal growth of pepper, such as the brown spot disease and the early blight disease. Meanwhile, we also collect healthy pepper leaves to expand our dataset. As shown in Table 2, there are 3921 pepper leaf images in our dataset including the healthy pepper leaves (HPL) and two different categories of infection (2606 images): spot disease (SD) and early blight disease

TABLE 2 Four datasets for the validating the proposed model on the pepper leaf.

Dataset	Test	Training	Validation	Total
Spot Disease (SD)	186	1015	184	1385
Early Blight Disease (EBD)	164	895	162	1221
Healthy Pepper Leaf (HPL)	176	965	174	1315
Total Pepper Leaf (TPL)	526	2875	520	3921

FIGURE 1
The sample dataset with different background.

(EBD), and several examples are shown in Figure 1. As shown in Table 2, the SD, EBD, and HPL datasets contain 1385, 1221, and 1315 images. The total pepper leaf (TPL) dataset is comprised of the SD, EBD, and HPL datasets. In our experiment, the images of each image dataset are split into the training set, the validation set and the test set, and the image numbers of the training set. Meanwhile, in order to evaluate the robustness of the BAF-Net, the images were taken with different complex background as shown in Figure 1.

2.2 Dataset labeling

In the following section, we present a data labeling process, and the labeled images are used for validating the proposed model. To accurately annotate the given images, we use the open-source tool named as LabelMe[1], which was developed by the computer science and artificial intelligence laboratory of MIT university. It allows users to annotate images manually to build image dataset for image segmentation. The pixel-by-pixel way carefully delineated the boundary of each leaf. All these images in the experiment are marked using this tool. Thereafter, each annotated image generates

a binary segmentation mask, where the intensity values of the foreground and background are 1 and 0, respectively. During annotating the dataset, we retain the same size as the input image. In view of the computational cost in deep learning, we set the size of the input image to 512×512.

2.3 Method

2.3.1 Overall architecture

In the field of image segmentation, U-Net has become one of the most successful network frameworks. It consists of a contracting path and an expanding path, where the contracting path is used to capture the image feature while the expanding path can achieve object localization. In each encoder-decoder layer, a skip connection layer transforms the low-level and high-level information. The model uses a convolution layer with fixed kernel size to extract image features, However, it is difficult to capture long-range semantic information. Although Transformers (Dong et al., 2019) can effectively encode the long-range dependencies, it is difficult to obtain local details and accurate boundaries of pepper leaves. To solve this problem, we propose a bidirectional attention fusion network by combining CNN and Transformer for pepper leaf segmentation, also named as BAF-Net,

---

1   https://github.com/wkentaro/labelme

where CNN is used to extract the local image information while the Transformer-based module can capture the long-range dependencies. As shown in Figure 2, the multi-scale branch is used to extract the global features while the full-scale feature can retain the detailed boundary information. The bidirectional fusion module is designed to concatenate the multi-scale features and the full scale features.

Specifically, BAF-Net includes three parts: a multi-scale feature fusion (MSFF) branch, a full-scale fusion feature (FSFF) branch, and bidirectional attention fusion (BAF) modules. In the MSFF branch, the network structure is similar to U-Net, composed of an encoding path and a decoding path. Different from the U-Net model, the encoder is replaced by a hybrid module by incorporating the convolutional layer and the Swin Transformer (Liu et al., 2021) module, and the decoder is composed of convolutional modules. In the FSFF branch, we first upsample four features: the output features of the encoder from the 2nd layer to the 4th layer, and the 5th layer of the decoder. Four output features are the same size as the first layer's output feature in the MSFF branch. Then, we fuse five generated features, and the generated feature is passed through four continuous convolutional modules. Each convolutional module is activated by the convolution layer, batch normalization, and the ReLU activation function. In the BAF module, the input features are from the output feature of the decoder in the MSFF branch and the output feature of the corresponding convolutional module in the FSFF branch. By incorporating the MSFF and FSFF branches, the improved model not only achieves the full resolution feature but also extracts the comprehensive and multi-scale features.

2.3.2 Multi-scale feature fusion branch

The transformer-based model (Dosovitskiy et al., 2020; Cao et al., 2021) has a more robust representation than the CNN-based model while building the long-range dependencies. In order to extract the global features, we explore a hybrid Swin-Trans-Conv block by combining the Swin-Transformer encoder and the convolutional layer, which is used to replace the convolutional layer of the encoder in the MSFF branch. As shown in Figure 3A, the backbone network including an encoder network and a decoder network is similar to U-Net. In the encoder network, we use a hybrid module by combining the convolutional layer and the Swin Transformer block, also called as Swin-Trans-Conv block, to replace each convolutional layer of U-Net, where an average pooling operator perform the downsampling process and the size of the feature maps are changed into half of the original. The decoder network is comprised of four convolutional layers and four upsampling operators. The upsampling operation is achieved by performing a deconvolutional operator with the stride of 2. The convolutional layer consists of a convolutional operator, batch normalization, and a ReLU activation layer. The number of channels in five layers corresponding to the 1st layer to the 5th layer is 32, 64, 128, 256 and 512, respectively.

Assuming that the input feature is $X \in R^{B \times H \times W \times C}$, where B, C, H and W represent the batch size, the channel number, and the image height and width of the input feature, respectively. In the Swin-Trans-Conv block as shown in Figure 3A, we first transforms the input feature X into $X' \in R^{B \times H \times W \times C}$. Then, we perform a 1×1 convolution operator on the generated feature, and split the generated feature Y into two groups F_{Trans} and F_{conv}, which can be expressed as:
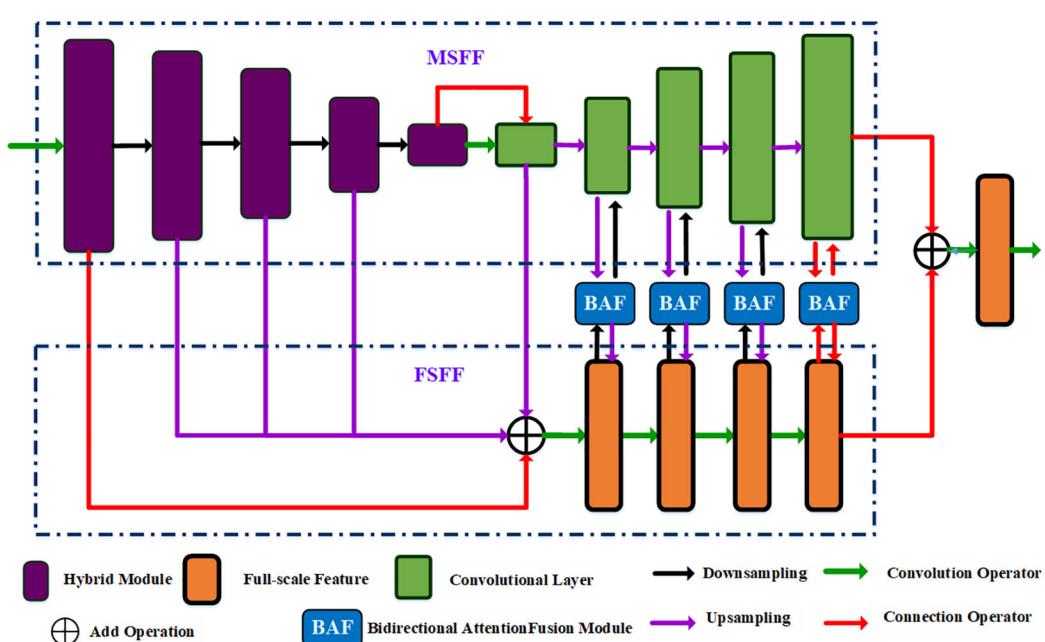

FIGURE 2
The overall framework of the proposed BAF-Net, which includes three main modules such as the multiscale feature fusion branch, GAM and decoders, where the decoder includes the global context module (GAM) and FAM with LAM.
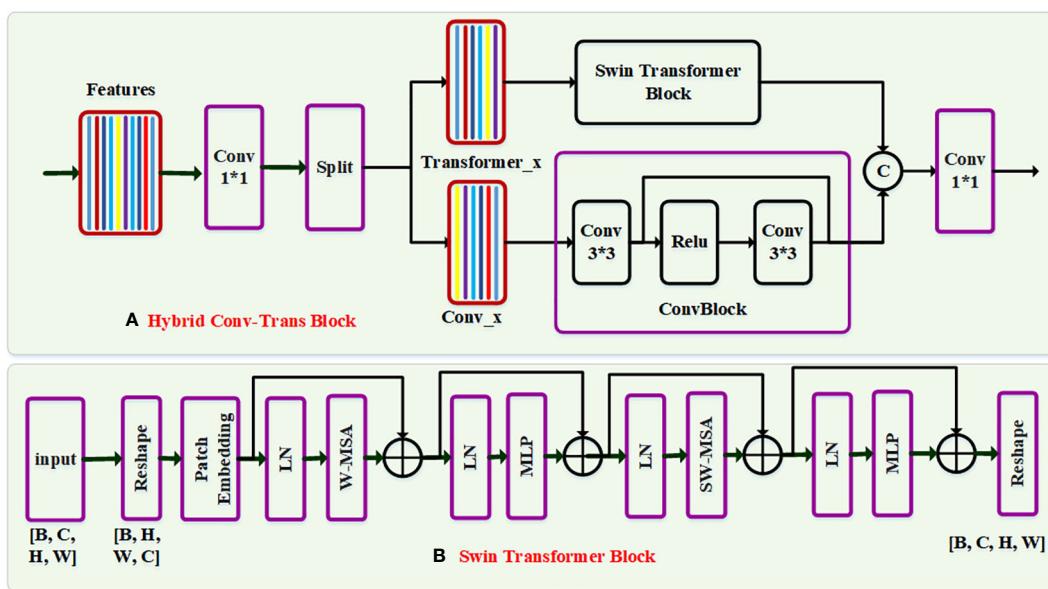
FIGURE 3

The network structure of the Swin-Trans-Conv block. In each block, the input feature is first passed through a 1×1 convolution, and subsequently is split evenly into two feature map groups, each of which is then fed into a Swin transformer block and a residual 3×3 convolutional (RConv) block, respectively. Afterwards, the output features of the Swin-Trans-Conv block and the RConv block are concatenated and then passed through a 1×1 convolution to generate a novel feature *via* a residual path.

$$Y = \text{Conv}_{1\times1}(\text{Re shape}(X)) \qquad X, Y$$
$$\in R^{B \times C \times H; \times W} \qquad (1)$$

$$F_{trans}, F_{conv} = \text{Split}(Y) \qquad F_{Trans}, F_{conv}$$
$$\in R^{B \times C/2 \times H \times W} \qquad (2)$$

where $\text{Re shape}(\cdot)$ is a reshape operator on two feature matrix F_{trans} and F_{conv}, $\text{Conv}_{1\times1}(\cdot)$ denotes a 1×1 convolutional operator, and $\text{Split}(\cdot)$ represents a split operation on the multidimensional matrix. Finally, the feature F_{Trans} is passed through a module based on Swin Transformer (Swin-Trans) encoder, and the generated feature map F'_{trans} is written as:

$$F'_{trans} = \text{SwinTrans}(F_{trans}) \qquad (3)$$

Similarly, the feature map F_{conv} passes through a residual convolution module, and the generated feature F'_{conv} is defined as:

$$F'_{conv} = \text{RConv}(F_{conv}) \qquad (4)$$

where $\text{RConv}(\cdot)$ is the residual convolution module, which is comprised of a 3×3 convolution filter, a ReLU activation layer, and a 3×3 convolution filter by a residual path, which is rewritten as:

$$F^3 = \text{Conv}_{3\times3}(F_{conv}) \qquad (5)$$

$$F'_{conv} = \text{Conv}_{3\times3}(\text{Re lu}(F^3)) + F^3 \qquad (6)$$

where F^3 is the feature map performed a 3×3 convolution operation on the feature map F_{conv}, $\text{Conv}_{3\times3}$ is a 3×3 convolutional layer, and $\text{Re lu}(\cdot)$ is a ReLU activation layer.

Finally, we concatenate two features F'_{trans} and F'_{conv}, and then perform a 1×1 convolution filter by a residual path, which is represented as:

$$X_{out} = \text{Conv}_{1\times1}(F'_{trans} \odot F'_{conv}) + X \qquad (7)$$

where $\odot$ denotes the concatenation operation.

Meanwhile, to construct the Swin-Trans module as shown in Figure 3B, the input feature F_{Trans} is split into small patches, and each patch size is set to $P \times P \times P$, where P is a positive integer and the number of the patches is $S=[H/P]\times[W/P]\times[C/P]$. For the feature F^i_{Trans} of the i-th layer with the 3D patches, we first compute the multi-head self-attention in a small window (W-MSA), which can be formulated as:

$$\hat{F}^i_{out} = \text{W-MSA}(\text{LN}(F^i_{Trans})) + F^i_{Trans} \qquad (8)$$

$$\hat{F}^i_{mlp} = \text{MLP}(\text{LN}(\hat{F}^i_{Out})) + \hat{F}^i_{Out} \qquad (9)$$

where $\text{W-MSA}(\cdot)$ denotes the window multi-head self-attention, $\text{LN}(\cdot)$ is the layer normalization operator, and $\text{MLP}(\cdot)$ denotes a multilayer perceptron module with two fully-connected layers and the GELU activation function. Then, the generated feature $\hat{F}^i_{mlp}$ is passed through the multi-head self-attention in the shifted window (SW-MSA), which is represented as:

$$\hat{F}^i_{sw} = \text{SW-MSA}(\text{LN}(\hat{F}^i_{mlp})) + \hat{F}^i_{mlp} \qquad (10)$$

$$F^i_{out} = \text{MLP}(\text{LN}(\hat{F}^i_{sw})) + \hat{F}^i_{sw} \qquad (11)$$

where $\text{SW-MSA}(\cdot)$ denotes the shifted window multi-head self-attention, and F^i_{out} is the output feature of the i-th layer.

Finally, the output feature F_{out}^i is reshaped into the same size of the input feature in the Swin-Trans module.

It is worth noting that the Swin-Trans-Conv block has several advantages. First, it integrates the local modeling capability of the convolution module and the global modeling capability of the Swin-Trans module. Secondly, the split and concatenation operations are used for two branches to extract different features, reducing the computational complexity and the number of parameters.

2.3.3 Full-scale feature fusion branch

The edge and detailed image information may be lost in the U-shape network framework due to the continuous downsampling operators. To solve this problem, we design an MSFF branch to retain the detailed information, and the network structure is shown in Figure 2B. We fuse the output features of the first 1st to 4th layer in the encoder of the MSFF branch and the output feature of the decoder of the 5th layer in the decoder since the multi-scale features can enhance the edge information (Liu et al., 2023). For four output features from the MSFF branch, we first carry out a 1×1 convolution filter to reduce the channel number. Then, we perform the upsampling operator on the four features, and the four generated features have the same size with the first channel feature. Then, we integrate four generated features into the input feature by a residual path, which can be expressed as:

$$X_i^{up} = \text{Conv}_{1\times1}\left(\underbrace{\text{up}(\cdots\text{up}(X_i))}_{i-1}\right) \qquad i = 2, \cdots, 5 \quad (12)$$

$$X_{fuse} = \sum_{i=2}^{5} X_i^{up} + X \qquad (13)$$

where $\text{Conv}_{3\times3}$ denotes a 3×3 convolutional filter.

Finally, the novel feature passes through four continuous convolutional modules. Each convolutional module includes a 3×3 convolution layer, batch normalization, and a ReLU activation layer. To reduce the computational cost and the parameters, we keep each channel number of four features equal to that of the first layer in the MSFF branch. In this paper, the channel number is set at 32. The operations for each convolution block are presented as follows:

$$\begin{cases} X_f^4 = \text{Re lu}(\text{BN}(\text{Conv}_{3\times3}(f_{fuse}))) & i = 5 \\ X_f^{i-1} = \text{Re lu}(\text{BN}(\text{Conv}_{3\times3}(X_f^i))) & i = 2, \cdots, 4 \end{cases}$$
$$(14)$$

2.3.4 Bidirectional attention fusion module

In order to achieve the multi-scale and full-scale features, we designed a BAF module to generate the corresponding output features for the MSFF and FSFF branches. As shown in Figure 4, the BAF module includes multi-scale feature guidance (MSFG) module and full-scale feature guidance (FSGM) module. For the MSGM module, we first conduct the downsampling operation on the input feature of the FSGM module, and the novel feature maps have the same spatial dimensions with the same with that of the MSGF map, which can be expressed as:

$$F_{fg}^{dn} = \text{DN}(F_{fg}) \qquad (15)$$

where $\text{DN}(\cdot)$ denotes the downsampling operation. Then, we concatenate the output feature of the MSFF branch F_{ms} and the feature F_{fg}^{dn}, and perform a 1×1 convolution module on the novel feature map to compress the number of channels, we can obtain the feature map:

$$F_{ms}^c = \text{Conv}_{1\times1}(F_{ms} \text{ e } F_{fg}^{dn}) \qquad (16)$$
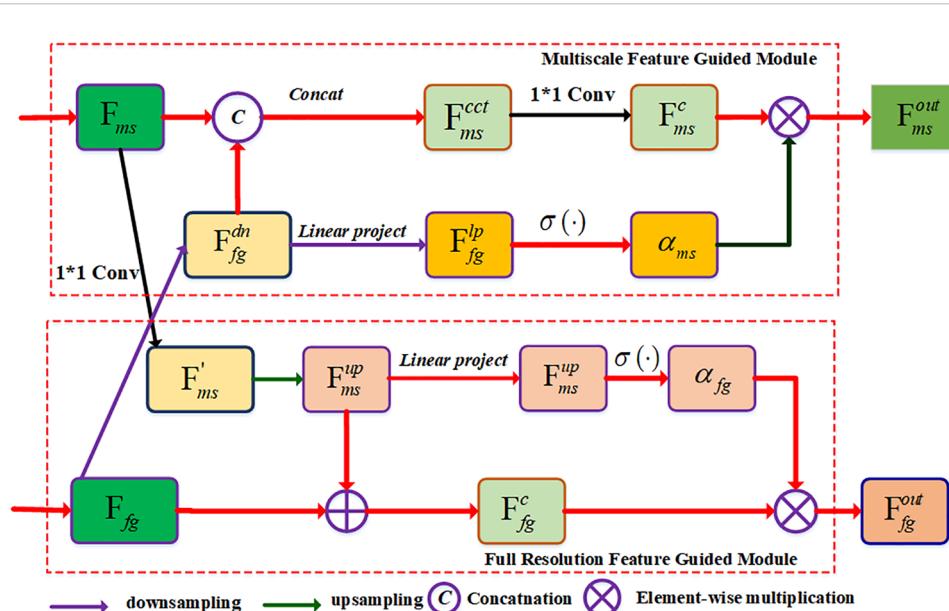
FIGURE 4

The network structure of the BAF module. Two input features F_{ms} and F_{fg} are from the output features of the MSFF and FSFF branches, respectively. The BAF module contains a multi-scale feature guided (MSFG) module and a full resolution feature guided (FRFG) module. The MSFG module is used to generate the multi-scale feature while the FRFG module is used to generate the full-scale feature.

At the same time, we project the feature F_{ms}^c to compress the feature map into a channel along the channel direction, and use the Sigmoid activation function to obtain the global attention map, which is defined as:

$$\alpha_{ms} = \sigma_{sig}(\text{Pr oj}(F_{ms}^c)) \tag{17}$$

where $\text{Pr oj}(\cdot)$ denotes the linear projection function, $\sigma_{sig}(\cdot)$ denotes the Sigmoid function, and $\alpha_{ms} \in [0,1]$ is the spatial attention map of the feature F_{ms}^c. It is obvious that the spatial attention map α_{ms} calculates the spatial weight of each pixel, and the calibrated feature map is expressed as:

$$F_{ms}^{out} = \alpha_{ms} \otimes F_{ms}^c \tag{18}$$

Finally, the feature F_{ms}^{out} is transformed to the next convolution layer.

In the FSFG module, we first perform a 1×1 convolutional filter on the multi-scale feature map F_{ms} to compress the number of channels. The expression is as follows:

$$F_{ms}^{'} = \text{Conv}_{1\times1}(F_{ms}) \tag{19}$$

Then, we upsample the multi-scale feature map to make the generated features have the same spatial dimension as that of the full-scale feature. The expression is as follows:

$$F_{ms}^{up} = \text{up}(F_{ms}^{'}) \tag{20}$$

where $\text{up}(\cdot)$ denotes the upsampling operator. Afterwards, two features F_{ms}^{up} and F_{fg} are fed into the convolutional layer to generate a new feature F_{fg}^c, which is written as:

$$F_{fg}^c = \text{Conv}_{3\times3}(F_{ms}^{up} \oplus F_{fg}) \tag{21}$$

where $\oplus$ represents the pixel-wise addition operation. Meanwhile, we use linear projection to compress the feature into a channel along the channel direction, and then use the sigmoid activation function to obtain the global attention map:

$$\alpha_{fg} = \sigma_{sig}(\text{Pr oj}(F_{fg}^c)) \tag{22}$$

where $\alpha_{fg} \in [0, 1]$ is the spatial attention map of F_{fg}^c, which is used to calculate the spatial position weight of each pixel. The calibrated feature map can be represented as:

$$F_{fg}^{out} = \alpha_{fg} \otimes F_{fg}^c \tag{23}$$

Finally, it is input to the convolution layer of the next FSFG module.

2.3.5 Training loss

The network should be trained to obtain the best training parameters. It is known that the loss function is essential to the predicted performance of the segmentation model. The loss function is used to measure the deviation between the model prediction and the ground truth. The binary cross entropy (BCE) is a loss function widely used in binary image segmentation tasks. Assuming that the input predicted result is p, and the corresponding ground truth label is g, the BCE loss function is defined as:

$$L_{bce}(p, \quad g) = -\sum_{i=1}^{N}[g_x \log(p_x) + (1-g_x)\log(1-p_x)] \tag{24}$$

The intersection over union (IoU) loss is defined as:

$$L_{IoU}(p, \quad g) = -\log\left(\frac{\sum_{i=1}^{N}|g_x \cdot p_x|}{\sum_{i=1}^{N}(g_x + p_x - |g_x \cdot p_x|)}\right) \tag{25}$$

Therefore, our final loss includes L_{bce} and L_{IoU}, which can be expressed as:

$$L_{total}(p, \quad g) = \alpha L_{bec}(p, g) + (1-\alpha)L_{IoU}(p, g) \tag{26}$$

The weight α is a coefficient to balance the importance of two loss functions, and we set $\alpha = 0.5$.

2.4 Performance evaluation

In order to verify the segmentation performance, we use six evaluation indices to evaluate the accuracy of the model on the pepper leaf datasets. Six evaluation indices include: pixel accuracy (PA), pixel recall (PR), pixel precision (PP), pixel specificity (PS), intersection over union (IoU) and F1 score. We assume that TP (True Positive) represents the number of pixels that are both 1 in the predicted value and the label value, TN (True Negative) represents the number of pixels that are both 0 in the predicted value and the label value, FP (False Positive) represents the number of pixels that are 1 in the predicted value and 0 in the label value, and FN (False Negative) represents the number of pixels that are 0 in the predicted value and 1 in the label value. The expression of the pixel accuracy is written as follows:

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \tag{27}$$

PR is defined as follows:

$$PR = \frac{TP}{TP + FN} \tag{28}$$

PP is defined as follows:

$$PP = \frac{TP}{TP + FP} \tag{29}$$

F1 score is defined as:

$$F1 = \frac{2 \times PR \cdot PP}{PR + PP} \tag{30}$$

PS is defined as follows:

$$PS = \frac{TN}{TN + FP} \tag{31}$$

From Equations (27)-(31) and the IoU as defined in Equation (25), it can be seen that six evaluation indices range from 0 to 1. The higher the index values are, the best segmentation performance is obtained. Generally speaking, the mean IoU (mIoU) is used to evaluate the segmentation performance on a given dataset.

3 Experiments

In this section, we present the experimental results including the experimental settings, the comparison with the state-of-the-arts models, the ablation study and the discussion.

3.1 Experimental settings

All models in the experiment are carried out on Intel (R) Core (R) i7-8700K CPU 3.70GHz CPU and Nvidia GeForce TITAN XP 12 GB GPU with 48G RAM. The programs are conducted on the Ubuntu 16.04 with the Conda environment. In the BAF-Net, the parameter settings are as follows: the batch size is set to 4, the number of iterations (epoch) is set to 60, and each epoch contains 350 batches. During the training process, the network is optimized using stochastic gradient descent (SGD), the initial learning rate is set to 0.01.

3.2 Comparison with the state-of-the-arts models

We compared BAF-Net with the state-of-the-art methods on four pepper leaf datasets, such as the SD, EBD, HPL, and TPL datasets. For fairness, these models are running on the same training dataset, the validation dataset, and the test dataset. The comparative models on the pepper leaf dataset involve U-Net (Ronneberger et al., 2015), AttU-Net (Oktay et al., 2018), Swin-UNet (Cao et al., 2021), SCUNet (Zhang et al., 2022) and the proposed BAF-Net. We set the training epochs to 60 for each trained model.

Table 3 shows the test results on the SD dataset using five different state-of-the-art models. Compared with U-Net, the proposed model has a precision increase of 7.48%, IoU increase of 3.88%, and F1 score increase of 5.0%. It also shows that PA score has the relative improvement of 0.5% on the SD dataset. For the attention U-Net model, the segmentation results on five indices are close to that of the U-Net. In addition, Swin-UNet and SCUNet have the similar segmentation performance. However, the segmentation performance of U-Net exceeds two models in terms of six evaluation indices. The reason is that Swin-UNet and SCUNet containing the transformer-based modules attain better segmentation results only if more efficient pre-trained model is provided. From Table 3, where the highest score for each indicator is shown in bold, our model can obtain the best segmentation

performance in terms of five evaluation indices including PA, PP, PS, IoU, and F1 scores compared with other models. By evaluating the segmentation performance of five models, we also give several examples of the segmentation results using these compared methods as shown in Figure 5.

Table 4 presents the segmentation results of five segmentation models on EBD pepper leaf dataset, in which the highest score of each index is shown in bold. From the experimental results, the proposed model has the highest scores among the six indices including PA, PR, PP, PS, IoU and F1 scores. Specifically, compared with U-Net, BAF-Net increased PA by 0.62%, PR by 0.06%, PP by 3.44%, PS by 1.7%, IoU by 5%, and F1 score by 7.04%. Attention U-Net is only lower than BAF-Net in terms of the indices IoU and F1-score, with a decrease of 4.92 and 6.59%, respectively. Compared with Swin-Unet and SCUNet, the proposed model has significant improvement in terms of six indices. The proposed BAF-Net have significant improvement in terms of PP, reaching the increase by 7.25% and 14.29%, respectively. By evaluating the segmentation performance of five deep learning-based models, we find these models can obtain better segmentation results than the traditional methods. Meanwhile, we also give the examples of the segmentation results using these compared methods as shown in Figure 6.

Table 5 shows the validation results of five different models on the HPL data set, with the highest score for each indicator shown in bold. From the experimental results, the proposed model can obtain the best segmentation accuracy in terms of PA, PP, PS, mIoU and F1 score. Compared with U-Net, the proposed model has increased PA by 0.21%, PP by 1.51%, PS by 1.52%, IoU by 0.01%, and F1 score by 0.27%. The attention U-Net has the similar segmentation results with U-Net. Our model has significant improvement than Swin-UNet and SCUNet in terms of the PP, mIoU and F1 score. Compared with the Swin-UNet, the PP, mIoU and F1 scores have increased by 4.95%, 3.60% and 2.43%, respectively. Compared with the SCUNet, the PP, mIoU and F1 score has increased by 2.51%, 2.22% and 1.58%, respectively. Meanwhile, we also give the example of the segmentation results for qualitative comparison, and the representative examples are shown in Figure 7.

The experimental results on the TPL dataset are shown in Table 6, with the highest score in each indicator represented in bold. It can be seen that our model obtains the best segmentation results in terms of the six indices among five models. Compared with U-Net, the proposed model has IoU increased by 0.01%, PA increased by 0.13%, PR increased by 0.03%, and PP increased by 0.87%. The PS score is 0.01% higher than that of U-Net, and F1 score is 0.48%

TABLE 3 The segmentation results on the SD dataset using five different models.

Model	PA(%)	PR(%)	PP(%)	PS(%)	mIoU(%)	F1(%)
UNet	98.70	**99.09**	91.10	98.65	91.80	94.93
Attention U-Net	98.24	98.11	88.72	98.26	90.09	93.18
Swin-UNet	97.68	97.90	85.36	97.65	86.76	91.20
SCUNet	97.42	98.14	83.68	97.32	87.34	90.33
Ours	**99.20**	98.74	**98.58**	**99.80**	**95.68**	**96.75**
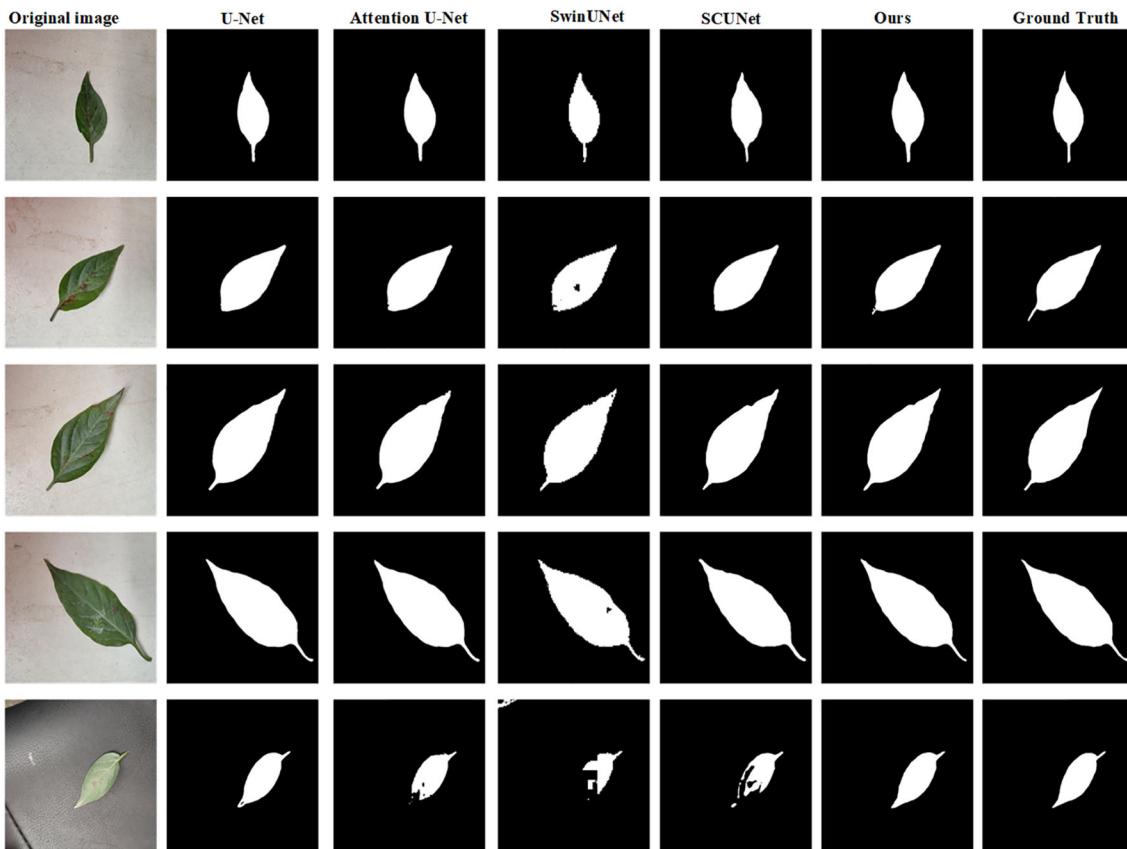
FIGURE 5
Examples of the predicted results using five different models on the SD dataset. From the 1st column to 7th column: the original images, the predicted results using the U-Net, attention U-Net, Swin-UNet, SCUNet, the proposed model and ground truth (GT), respectively.

higher than that of U-Net. Compared with attention U-Net, the proposed model has significant improvement in terms of six indices. However, Swin-UNet and SCUNet do not improve the segmentation results compared with the U-Net. In summary, BAF-Net has obvious advantages in segmenting the pepper leaf from the natural images.

3.3 Ablation study

In this section, we perform an ablation study to validate the effectiveness of each module. Especially, we consider the basic U-Net architecture as the baseline, namely the simple U-Net (SU-Net), which is similar to U-Net with half of the channel number of U-Net.

In the ablation experiments, we take SU-Net, MFF, MRF, and BAF as four basic modules. Our experimental strategy is to add a module each time, and it is proven to be effective. We approve that it is effective in subsequent studies. Strictly speaking, we selected four unique models, such as SU-Net, SU-Net-MFF, SU-Net- MFF-MRF, and BAF-Net, to verify that different modules are still valid when each model is added to SU-Net each time.

As shown in Table 7, we first experiment SU-Net-MSFF by replacing the convolution layer of the encoder in the SU-Net model with the Swin-Trans-Conv block, which is formulated by adding the MSFF module into SU-Net. Experiments show that PA, PR, PP, PS, mIoU and F1 score of the SU-Net-MSFF model are 98.94%, 96.87%, 96.90%, 99.36%, 95.63% and 96.88, respectively. Then, by adding the FSFF module to SU-Net-MSFF, the results show that the PA,

TABLE 4 The segmentation results on the EBD dataset using different models.

| Model	PA(%)	PR(%)	PP(%)	PS(%)	mIoU(%)	F1 (%)
UNet	96.98	98.11	81.54	95.82	81.76	84.06
Attention U-Net	95.57	96.63	75.08	95.41	81.84	84.51
Swin-Unet	96.02	95.55	77.73	96.08	80.63	85.72
SCUNet	94.49	95.70	70.69	94.32	77.33	81.31
BAF-Net(ours)	**97.60**	**98.17**	**84.98**	**97.52**	**86.76**	**91.10**
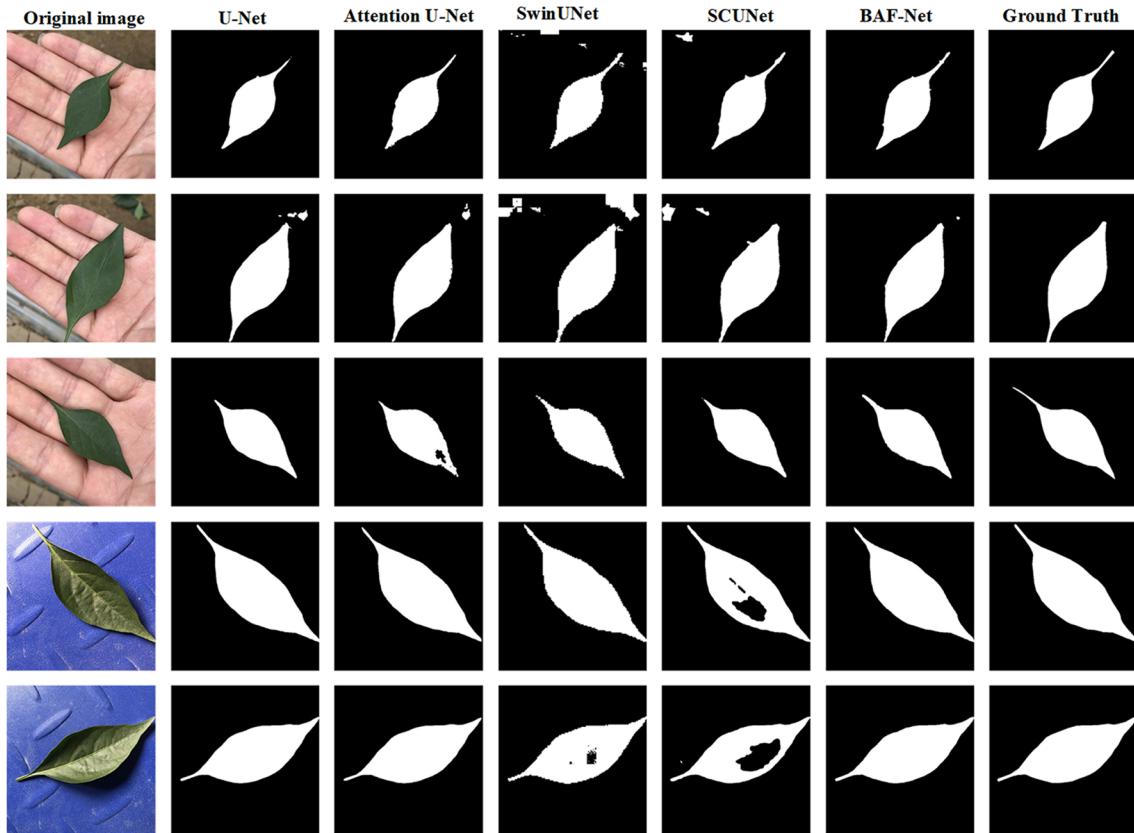
FIGURE 6
Examples of the predicted results using five different model on the EBD dataset. From the 1st column to 7th column: the original image, the predicted results using the U-Net, attention U-Net, Swin-UNet, SCUNet, and our model, respectively.

PA, PP, PS, IoU and F1 scores of the SU-Net-MSFF-FSFF are 98.94%, 96.62%, 97.14%, 99.42%, 95.68% and 96.98%, respectively. Compared with the SU-Net-MSFF, PR, PS, IoU and F1 score of the SU-Net-MSFF-FSFF model are increased by 0.24%, 0.06%, 0.05% and 0.10%, respectively. Finally, we experiment BAF-Net by fusing the output features of the decoder in the MSFF and FSFF branches to the BAF modules. The results show that PA, PR, PP, PS, IoU and F1 score of BAF-Net are 98.98%, 6.82%, 97.20%, 99.43%, 95.86% and 97.01%, respectively, which are increased by 0.04%, 0.2%, 0.06%, 0.01%, 0.18% and 0.03%, respectively. From the segmentation results, we can see that the addition of the Swin-Trans-Conv block expands the receptive field and enhances the feature extraction ability of SU-Net, enabling it to obtain different levels of information at the same time. The full-resolution features enable the proposed model to retain image local details. By combining multi-scale information and full scale information, it can extract deeper structural information. Therefore, the combination of the three modules can obtain the best performance.

3.4 Discussion

The above analysis shows that the segmentation results of these deep learning-based segmentation models are suitable. Compared with the classical methods based on the variational statistics theory (Costa et al., 2019; Fang et al., 2019a; Fang et al., 2019b; Gao and Lin, 2019; Liu et al., 2020; Fang et al., 2021a; Fang et al., 2021b; Liu et al., 2021; Ward et al., 2021), the deep-learning-based models can obviously obtain

TABLE 5 The segmentation results on the HPL dataset using different models.

| Model	PA(%)	PR(%)	PP(%)	PS(%)	mIoU(%)	F1 (%)
U-Net	99.09	**98.51**	95.84	99.02	96.11	97.07
Attention U-Net	99.21	97.69	96.34	99.44	95.53	97.01
Swin-UNet	98.63	97.57	92.40	98.79	92.52	94.91
SCUNet	98.88	96.69	94.84	99.21	93.90	95.76
BAF-Net(ours)	**99.30**	97.32	**97.35**	**99.60**	**96.12**	**97.34**
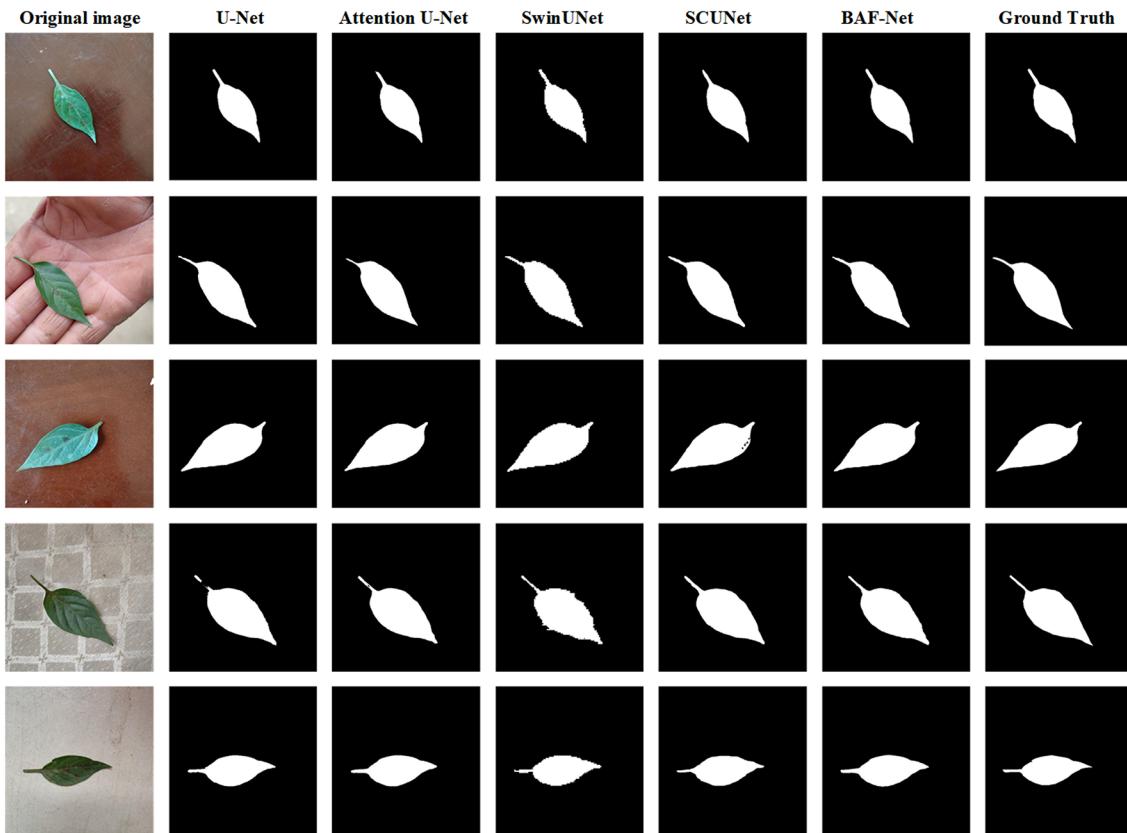
FIGURE 7
Examples of the predicted results using five different model on the HPL dataset. From the 1ˢᵗ column to 7ᵗʰ column: the original image, the predicted results using the U-Net, attention U-Net, Swin-UNet, SCUNet, and our model, respectively.

better classification results. In our work, to capture the long-range dependencies between different pixels, we propose a bidirectional adaptive attention fusion network called BAF-Net by exploring an adaptive attention mechanism to extract multi-scale and full-scale features simultaneously. Specifically, we first design an MSFF branch based on the encoder-decoder structure, which can not only extract local information of the target, but also learn the spatial attention to increase the receptive field. To further retain the boundary information of the segmented object, we propose a FSFF branch, and design adaptive bidirectional attention modules to achieve the bidirectional connection between the MSFF module and the FSFF module.

The results of the ablation experiment in Table 7 shows that progressive network such as SU-Net, SU-Net-MSFF, SU-Net-MSFF-FSFF and BAF-Net can improve the predicted performance of the baseline (SU-Net). Compared with the baseline, three models by progressively adding the MSFF, FSFF and BAF modules increase mIoU by 0.28%, 0.33% and 0.51%, respectively, and F1 score increased by 0.02%, 0.02% and 0.36%, respectively. From the segmentation results, it can be seen that BAF-Net has achieved the best performance. Compared with the baseline, the mIoU and F1 score of BAF-Net reaches 95.86% and 97.01%, respectively.

Although the proposed BAF-Net can obtain better performance on the four pepper leaf datasets, there are disadvantages in this work. (1) In the training process, the epoch number in our model is set to 60. Therefore, we need explore a schema to stop the training process for the deep learning-based model automatically. (2) Our

TABLE 6 The segmentation results on the pepper leaf dataset using different models.

Model	PA(%)	PR(%)	PP(%)	PS(%)	mIoU(%)	F1 (%)
U-Net	98.40	98.59	89.70	98.37	91.43	93.94
Attention U-Net	97.73	97.51	86.28	97.76	89.31	91.55
Swin-UNet	97.48	97.06	85.07	97.54	86.87	90.67
SCUNet	97.00	96.88	82.40	97.01	86.32	89.06
BAF-Net(ours)	**98.53**	**98.62**	**90.57**	**98.52**	**91.44**	**94.42**

TABLE 7 Comparison of pepper segmentation results of four models on the dataset.

Model	MSFF	FSFF	BAF	PA%	PR%	PP%	PS%	IoU%	F1%
SU-Net				98.86	96.52	96.80	99.35	95.35	96.66
SU-Net-MSFF	√			98.94	**96.87**	96.90	99.36	95.63	96.88
SU-Net_MSFF-FSFF	√	√		98.94	96.62	97.14	99.42	95.68	96.98
BAF-Net	√	√	√	**98.98**	96.82	**97.20**	**99.43**	**95.86**	**97.01**

model is supervised learning, which requires many training samples. Accordingly, in our future work, we will focus on the semi-supervised or self-supervised segmentation methods to reduce the requirements for training samples.

4 Conclusion

In our work, we propose a bidirectional adaptive attention fusion network for automatic segmentation of pepper leaves. The proposed model consists of the MSFF branch with the like-U-Net network structure, the FSFF branch, and the BAF modules with an adaptive attention mechanism. This MSFF branch fuses the Swin-Transformer-based and CNN-based modules to construct the Swin-Trans-Conv block, which replaces the convolution layer of the encoder of U-Net to expand the receptive field. In the MSFF branch, the CNN-based layer can extract the local image features while the Swin-Transformer-based module is used to extract the long-range dependencies of the channel and spatial information to expand receptive field. The FSFF branch performs multiple convolution layers keeping the same size with the original image, which is used to retain the boundary information and detail information of the segmented object. In addition, the BAF modules are used to fuse the output features of the MSFF and FSFF branch, which output the corresponding features for each branch. Compared with the existing model, our model obtain the highest evaluation indices on four pepper leaf datasets. In addition, the ablation experiment shows that the proposed three modules including MSFF, FSFF and BAF are effective. In the future, we will explore a weak-supervised model for pepper leaf segmentation since the small dataset may cause over-segmentation. Meanwhile, we study the construction of loss function and the method for augmentation dataset.

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

JF: conceptualization, methodology, experiment, and writing. HL: supervision and writing- review & editing. HJ: experiment. YF: methodology and investigation. SZ: methodology and investigation. LS: experiment. XH: writing-review & editing. JL: review & editing. MG: review & editing. All authors contributed to the article and approved the submitted version.

Funding

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

Aich, S., and Stavness, I. (2017). "Leaf counting with deep convolutional and deconvolutional networks," in *Proceedings of the IEEE international conference on computer vision workshops*, 2080–2089. doi: 10.1109/ICCVW.2017.244

Amean, Z. M., Low, T., and Hancock, N. (2021). Automatic leaf segmentation and overlapping leaf separation using stereo vision. *Array* 12, 100099. doi: 10.1016/j.array.2021.100099

Bhagat, S., Kokare, M., Haswani, V., Hambarde, P., and Kamble, R. (2021). "Wheatnet-lite: a novel light weight network for wheat head detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 1332–1341. doi: 10.1109/ICCVW54120.2021.00154

Bhagat, S., Kokare, M., Haswani, V., Hambarde., P., and Kamble, R. (2022). Eff-UNet ++: A novel architecture for plant leaf segmentation and counting. *Ecol. Inf.* 68, 101583. doi: 10.1016/j.ecoinf.2022.101583

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., et al. (2021). Swin-unet: Unet-like pure transformer for medical image segmentation. In: *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III* Cham: Springer Nature Switzerland, 205–218. doi: 10.48550/arXiv.2105.05537

Chandra, A. L., Desai, S. V., Guo, W., and Balasubramanian, V. N. (2020). Computer vision with deep learning for plant phenotyping in agriculture: a survey. *arXiv.* doi: 10.48550/arXiv.2006.11391

Costa, C., Schurr, U., Loreto, F., Menesatti, P., and Carpentier, S. (2019). Plant phenotyping research trends, a science mapping approach. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01933

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., et al. (2019). "Unified language model pre-training for natural language understanding and generation", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 13063–13075. doi: 10.48550/arXiv.1905.03197

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner,, et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv.* doi: 10.48550/arXiv.2010.11929

Fang, J., Liu, H., Zhang, L., Liu, J., and Liu, H. (2019a). Fuzzy region-based active contours driven by weighting global and local fitting energy. *IEEE Acess* 7, 184518–184536. doi: 10.1109/ACCESS.2019.2909981

Fang, J., Liu, H., Zhang, L., Liu, J., and Liu, H. (2019b). Active contour driven by weighted hybrid signed pressure force for image segmentation. *IEEE Acess* 7, 97492–97504. doi: 10.1109/ACCESS.2019.2929659

Fang, J., Liu, H., Zhang, L., Liu, J., and Liu, H. (2021a). Region-edge-based active contours driven by hybrid and local fuzzy region-based energy for image segmentation. *Inf. Sci.* 546 (6), 397–419. doi: 10.1016/j.ins.2020.08.078

Fang, J., Liu, H., Zhang, L., Liu, J., Zhou, H., and Liu, H. (2021b). Fuzzy region-based active contour driven by global and local fitting energy for image segmentation. *Appl. Soft Comp.* 100, 106982:1–16. doi: 10.1016/j.asoc.2020.106982

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

Hasan, A. S. M. M., Sohel, F., Diepeveen, D., Laga, H., and Jones, M. G. K. (2021). A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* 184, 106067. doi: 10.1016/j.compag.2021.106067

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 4700–4708. doi: 10.48550/arXiv.1608.06993

Koirala, A., Walsh, K. B., Wang, Z., and McCarthy, C. (2019). Deep learning-method overview and review of use for fruit detection and yield estimation. *Comp. Elec. Agric.* 162, 219–234. doi: 10.1016/j.compag.2019.04.017

Kumar, P., and Domnic, S. (2019). Image based leaf segmentation and counting in rosette plants. *Inf. Proce. Agric.* 6 (2), 233–246. doi: 10.1016/j.inpa.2018.09.005

Kuznichov, D., Zvirin, A., Honen, Y., and Kimmel, R. (2019). "Data augmentation for leaf segmentation and counting tasks in rosette plants,", *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* Long Beach, CA, USA, 2019, pp. 2580–2589. doi: 10.1109/CVPRW.2019.00314

Larese, M. G., Namías, R., Craviotto, R. M., Arango, M. R., Gallo, C., and Granitto, P. M. (2014). Automatic classification of legumes using leaf vein image features. *Patt. Recog.* 47, 158–168. doi: 10.1016/j.patcog.2013.06.012

Li, Z., Pan, J., Wu, H., Wen, Z., and Qin, J. (2020). "Memory-efficient automatic kidney and tumor segmentation based on non-local context guided 3D U-net," in *International conference on medical image computing and computer-assisted intervention (MICCAI)* (Springer), 197–206. doi: 10.1007/978-3-030-59719-1_20

Lin, K., Gong, L., Huang, Y., Liu, C., and Pan, J. (2019). Deep learning-based segmentation and quantifcation of cucumber powdery mildew using convolutional neural network. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00155

Lin, X., Li, C. T., Adams, S., Kouzani, A., Jiang, R., He, L., et al. (2023). Self-supervised leaf segmentation under complex lighting conditions. *Patt. Recog* 135, 109021. doi: 10.1016/j.patcog.2022.109021

Liu, H., Fang, J., Zhang, Z., and Lin, Y. (2020). A novel active contour model guided by global and local signed energy-based pressure force. *IEEE Access* 8, 59412–59426. doi: 10.1109/ACCESS.2020.2981596

Liu, H., Fang, J., Zhang, Z., and Lin, Y. (2021). Localised edge-region-based active contour for medical image segmentation. *IET Image Process.* 15 (7), 1567–1582. doi: 10.1049/ipr2.12126

Liu, H., Fu, Y., Zhang, S., Liu, J., and Fang, J. (2022). Active contour driven by adaptive-scale local-energy signed pressure force function based on bias correction for medical image segmentation. *IET Image Process.* 16 (14), 3929–3947. doi: 10.1049/ipr2.12604

Liu, H., Fu, Y., Zhang, S., Liu, J., Wang, Y., and Wang, G. (2023). GCHA-net: Global context and hybrid attention network for automatic liver segmentation. *Comput. Biol. Med.* 152, 106352. doi: 10.1016/j.compbiomed.2022.106352

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022). In *Proceedings of the IEEE/CVF international conference on computer vision. arXiv*, 10012–10022. doi: 10.48550/arXiv.2103.14030

Milletari, F., Navab, N., and Ahmadi, S. A. (2016). "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)* (IEEE). doi: 10.48550/arXiv.1606.04797

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). "Attention U-Net: Learning Where to Look for the Pancreas." *arXiv* abs/1804.03999. doi: 10.48550/arXiv.1804.03999

Ramachandran, P., Parmar, N., Vaswani, A., Bello, I., Levskaya, A., and Shlens, J. (2019). *Stand-alone self-attention in vision models* (NIPS). doi: 10.48550/arXiv.1906.05909

Ren, M., and Zemel, R. S. (2017). "End-to-end instance segmentation with recurrent attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 6656–6664. doi: 10.1109/CVPR.2017.39

Romera-Paredes, B., and Torr, P. H. S. (2016). "Recurrent instance segmentation," in *European Conference on computer vision* (Springer), 312–329. doi: 10.48550/arXiv.1511.08250

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention*, 9351 234–241 . doi: 10.1007/978-3-319-24574-4_28

Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., et al. (2016). Leaf segmentation in plant phenotyping: a collation study. *Mach. Vis. Appl.* 27 (4), 585–606. doi: 10.1007/s00138-015-0737-3

Shen, D., Wu, G., and Suk, H. I. (2017). Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* 19, 221. doi: 10.1146/annurev-bioeng-071516-044442

Singh, V. (2019). Sunflower leaf diseases detection using image segmentation based on particle swarm optimization. *Artif. Intel. Agric.* 3, 62–68. doi: 10.1016/j.aiia.2019.09.002

Slaughter, D. C., Giles, D., and Downey, D. (2008). Autonomous robotic weed control systems: A review. *Comp. Elec. Agric.* 61, 63–78. doi: 10.1016/j.compag.2007.05.008

Tian, K., Li, J., Zeng, J., Evans, A., and Zhang, L. (2019). Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm. *Comp. Elec. Agric.* 165, 104962. doi: 10.1016/j.compag.2019.104962

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01190

Wang, Z., Wang, K., Yang, F., Pan, S., and Han, Y. (2018). Image segmentation of overlapping leaves based on chan-vese model and sobel operator. *Inform. Proc. Agric.* 5 (1), 1–10. doi: 10.1016/j.inpa.2017.09.005

Ward, D., Moghadam, P., and Hudson, N. (2021). Deep leaf segmentation using synthetic data. *arXiv.* doi: 10.48550/arXiv.1807.10931

Xia, C., Lee, J. M., Li, Y., Song, Y. H., Chung, B. K., and Chon, T. S. (2013). Plant leaf detection using modified active shape models. *Biosyst. Eng.* 116, 23–35. doi: 10.1016/j.biosystemseng.2013.06.003

Zhang, K., Li, Y., Liang, J., Cao, J., Zhang, Y., Tang, H., et al. (2022). Practical blind denoising *via* swin-Conv-UNet and data synthesis. *arXiv.* doi: 10.48550/arXiv.2203.13278

Zhang, Z., Liu, Q., and Wang, Y. (2018). Road extraction by deep residual U-net. *IEEE Geosci. Remote Sens. Lett.* 15 (5), 749–753. doi: 10.1109/LGRS.2018.2802944

Zhou, Z., Siddiquee , M. M. R., Tajbakhsh , N., and Liang , J. (2019). UNet++: redesigning skip connections to exploit multiscale features in image segmentation . *IEEE Trans. Med. Imaging* 39 (6), 1856-1867. doi: 10.1109/TMI.2019.2959609

Check for updates

Grasping and cutting points detection method for the harvesting of dome-type planted pumpkin using transformer network-based instance segmentation architecture

Jin Yan, Yong Liu*, Deshuai Zheng and Tao Xue

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

An accurate and robust keypoint detection method is vital for autonomous harvesting systems. This paper proposed a dome-type planted pumpkin autonomous harvesting framework with keypoint (grasping and cutting points) detection method using instance segmentation architecture. To address the overlapping problem in agricultural environment and improve the segmenting precision, we proposed a pumpkin fruit and stem instance segmentation architecture by fusing transformer and point rendering. A transformer network is utilized as the architecture backbone to achieve a higher segmentation precision and point rendering is applied so that finer masks can be acquired especially at the boundary of overlapping areas. In addition, our keypoint detection algorithm can model the relationships among the fruit and stem instances as well as estimate grasping and cutting keypoints. To validate the effectiveness of our method, we created a pumpkin image dataset with manually annotated labels. Based on the dataset, we have carried out plenty of experiments on instance segmentation and keypoint detection. Pumpkin fruit and stem instance segmentation results show that the proposed method reaches the mask mAP of 70.8% and box mAP of 72.0%, which brings 4.9% and 2.5% gains over the state-of-the-art instance segmentation methods such as Cascade Mask R-CNN. Ablation study proves the effectiveness of each improved module in the instance segmentation architecture. Keypoint estimation results indicate that our method has a promising application prospect in fruit picking tasks.

KEYWORDS

keypoint detection, stem instance segmentation, transformer, point rendering, pumpkin harvesting

1 Introduction

Agriculture is the foundation of people's livelihood. To effectively harvest crops, fruits and vegetables, researchers have made efforts from different aspects, for instance, nutrient supply (Sun et al., 2022), disease prevention (Yang et al., 2022), postharvest preservation (Pan et al., 2023) and so on. Crop, fruit and vegetable picking is often the most labor-intensive part of the entire production chain. Therefore, intelligent picking robots have become a research hotspot. Among them, accurate detection is a prerequisite for intelligent picking, and many vision-based fruit and vegetable detection works have been launched.

In recent years, deep learning applications have attracted great attention and made great breakthroughs in image processing tasks (Liu et al., 2021a; Bhatti et al., 2023), the research on learning-based fruit and vegetable detection also moves forward. Liu et al. (2019a) trained a Support Vector Machine (SVM) classifier utilizing the Histograms of Oriented Gradients (HOG) descriptor to detect mature tomatoes. The proposed machine learning method's recall, precision, and F1 scores are 90.00%, 94.41%, and 92.15%, respectively. Sun et al. (2019) designed a GrabCut model based on the visual attention mechanism for fruit region extraction, then applied the Ncut algorithm to segment the extracted fruits. The recognition method achieves the F1 score of 94.12% and an error of 7.37%. Deep learning (DL) has developed rapidly in these years, and because of its excellent performance, DL has been applied in many fields, including agriculture. Yuan et al. (2020) applied SSD to detect tomatoes in the greenhouse with the backbone of Inception V2, and the network achieves an average precision of 98.85%. Bresilla et al. (2019) set up a fruit detection network based on YOLO. The network can be trained to detect apples and pears without classifying them. The architecture shows an accuracy of more than 90% fruit detection. Fu et al. (2020) compared two Faster R-CNN based architectures ZFNet and VGG16, employed to detect apples in images. The results indicate that the network with VGG16 achieves the highest average precision (AP) of 0.893.

It can be seen that the accuracy and speed of fruit and vegetable detection can meet the requirements of practical applications. However, deep learning-based detection frameworks only generate coarse boundaries, and many pixels irrelevant to the detected fruit or vegetable are also included in the bounding box. In order to obtain more abundant information, some scholars have carried out researches on fruit or vegetable instance segmentation. Instance segmentation combines the advantages of semantic segmentation and object detection and identifies each object instance of each pixel for every known object within an image. With the help of instance segmentation, fruits or vegetables can be assigned to different instances with pixel accuracy.

Ganesh et al. (2019) presented a deep learning approach, named Deep Orange, to detect and pixel-wise segment oranges based on Mask R-CNN. Gonzalez et al. (2019) proposed a network based on Mask R-CNN for blueberry detection and instance segmentation. The authors tested the performances of several backbones such as ResNet101, ResNet50, and MobileNetV1. Jia et al. (2020) improved Mask R-CNN through the fact as the feature extraction, RoI acquisition, and mask generation so that the network is more

suitable for recognizing and segmenting overlapped apples. Also based on the well-known Mask R-CNN network, Perez-Borrero et al. (2020) designed a new backbone and mask network, removed the object classifier and the bounding-box regressor and replaced the non-maximum suppression algorithm with a new region grouping and filtering algorithm to better segment instances of strawberry. The same research team (Perez-Borrero et al., 2021) proposed another strawberry instance segmentation methodology based on the use of a fully convolutional neural network. Instance segmentation is achieved by adding two new channels to the network output so that each strawberry pixel predicts the centroid of its strawberry. The final segmentation of each strawberry is obtained by applying a grouping and filtering algorithm. Liu et al. (2019b) improved Mask R-CNN to detect and segment cucumbers by designing a logical green operator to filter non-green backgrounds and adjusting the scales and aspect ratios of anchor boxes to fit the size and shape of cucumbers.

In the actual picking applications, the key operating points are often generated in the fruit stem area, so the detection of stems should be taken seriously. Some scholars have focused their attention on fruit stem detection.

Sa et al. (2017) made use of an RGB-D sensor to acquire color and geometry information and utilized a supervised-learning approach for the peduncle detection task. Yoshida et al. (2018) used the support vector machine to classify the point cloud data, clustering to obtain fruit stem pixels, and then looking for cutting points. Luo et al. (2018) studied the detection of cutting points on stems of overlapping grape clusters. After segmenting individual clusters using machine learning method, a geometric constraint method is then used to determine the cutting point in the region of interest of each cluster's stem. Sun et al. (2021) developed a deep learning-based top-down framework to detect keypoint on the bearing branch, enabling branch pruning during fruit picking. This work only detects citrus branch keypoint without segmentation. Kalampokas et al. (2021) applied a regression convolutional neural network (RegCNN) for executing a stem segmentation task and determined the cutting point on the stem based on a geometric model. Chen et al. (2021) proposed a banana stalk segmentation method based on a lightweight multi-feature fusion deep neural network. The methods in both (Kalampokas et al., 2021) and (Chen et al., 2021) can only segment the stem of a single cluster of grape or banana. Wan et al. (2022) proposed a real-time branch detection and reconstruction method applied to fruit harvesting. To segment the branches separately, the authors first detect branch region boxes using YOLOv4, then utilize image segmentation to locate the branch boundaries. Next, the division of precise boxes belonging to the same branch is achieved based on the branch growth trend constraints. Rong et al. (2021) proposed a method to localize the peduncle cutting point and estimate the cutting pose. The authors first detect tomatoes *via* YOLOv4 and then segment fruit and peduncle masks by YOLACT++. The segmented peduncle mask is fitted to the curve using least squares and three key points on the curve are found. Chen and Chen (2020) proposed a methodology to identify the plucking points of tea shoots using machine vision and deep learning. The authors first localize the one tip with two leaves regions through Faster-RCNN,

then identify the plucking areas using FCN. The plucking point is determined as the centroid of the plucking area. The approaches in (Wan et al., 2022) (Rong et al., 2021), and (Chen and Chen, 2020) treat detection and stem instance segmentation as two separate networks.

As a nutritious crop, there are few studies on pumpkin detection. Wittstruck et al. (2020) and Midtiby and Pastucha (2022) have conducted researches on large-scale pumpkin yield estimation. The datasets are captured by UAVs from the air. To the best of the authors' knowledge, there is currently no dataset consisting of close-range pumpkin images and devoted to autonomous pumpkin harvesting. In this paper, we established a dataset on two varieties of pumpkin, and the instance masks of pumpkin fruit and pumpkin stem are labeled manually. The pumpkin stem is thick and it is hard to tear off or twist off the pumpkin fruit with one end effector. As is illustrated in Figure 1B, an ideal way to pick the pumpkin is utilizing two arms or one arm with two end effectors, one to grasp and another to cut. The detection of pumpkin stems cannot be ignored during automatic picking. This paper presents a pumpkin autonomous picking framework with keypoint detection and instance segmentation method. Firstly, pumpkin fruit and stem masks can be generated by instance segmentation method as shown in Figure 1A. Then, through the keypoint detection algorithm, relationships among the fruit and stem instances are determined and keypoints are localized as marked in Figure 1, where red points are cutting points, blue points are grasping points, and yellow lines link one stem and one fruit that belong to one pumpkin instance. Main contributions of our work are three folds:

1) We propose a novel pumpkin autonomous picking framework with grasping and cutting point detection method using instance segmentation architecture. The keypoint detection algorithm can model the relationships among the fruit and stem instances as well as estimate grasping and cutting keypoints.

2) This paper presents a pumpkin fruit and stem instance segmentation architecture based on deep learning and applying a transformer backbone and point rendering mask head. Compared with several state-of-the-art instance segmentation methods, the proposed method shows significant performance advantages in both metric evaluation and visualization analysis.

3) To validate the effectiveness of our method, we created a pumpkin image dataset with manually annotated labels. Downstream tasks such as image classification, pumpkin detection and instance segmentation can be deployed on the database.

The remainder of this paper is arranged as follows. Section 2 introduces the dataset and our method. Section 3 presents the results and analyses. Finally, conclusions are summarized in Section 4.

2 Materials and methods

In this paper, we perform instance segmentation on pumpkin fruit and stem. Then, we detect and localize the grasping points and cutting point using the proposed keypoint detection algorithm. To complete this research, we first collect pumpkin images to establish the dataset.

2.1 Data acquisition

This paper establishes a pumpkin dataset containing two varieties of pumpkin (Bebe pumpkin and Hazel pumpkin). The dataset was collected in Tangshancuigu modern agriculture demonstration zone, Nanjing, China. We used three different capture devices (Intel RealSense D435i, One Plus 6T smartphone, and Apple iPhone 13 Pro smartphone) to collect a total of 679 ripe pumpkin images. The original image pixels are 1280×720, 4608×3456, and 4032×3024, respectively. To better train the images, we resized the high-resolution images from 4608×3456 and 4032×3024 to 640×480. The resolutions of final images in the dataset are 1280×720 and 640×480. The dataset collection environment and real image examples are shown in Figure 2.
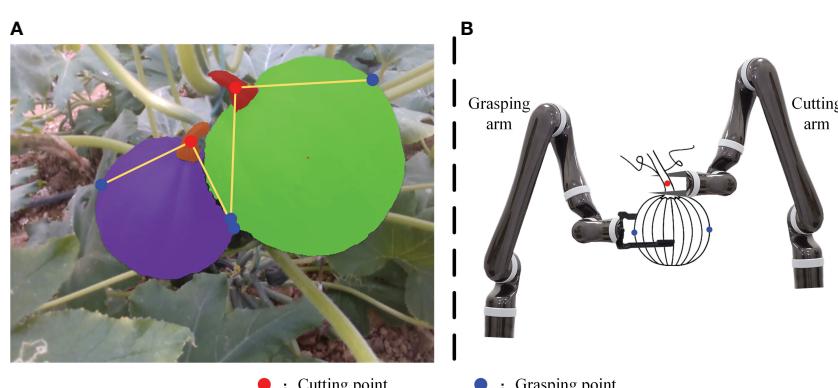

FIGURE 1
Example of pumpkin keypoint detection framework output **(A)** and pumpkin picking illustration **(B)**.

Through the review above, the pixel-wise segmentation of fruits and the labeling of fruit stems are very important. Therefore, we manually annotated the pixel-level instances of the pumpkin fruit and stem, as well as the pumpkin box containing one fruit and stem (see Figure 3). The labeling software we used is Labelme. Table 1 shows the distribution of the dataset.

The data is split into a training set and a testing set with a ratio of 80:20, where 543 images belong to training set and 136 images are in the testing set. Differing from the general structured scene, agricultural environment is a typical unstructured scene. The key problems faced during image collection in agricultural environment are large changes in illumination, a lot of dust, and frequent overlaps of fruit branches and leaves. To simulate the agriculture environment and enhance the generalization and robustness of deep neural network, we augment the dataset by changing brightness, blurring the image, adding noise, and cutout operation as shown in Figure 4. In addition, the horizontal flip is operated with a probability of 0.5 during training. After data augmentation, the training set contains 3258 images.

2.2 Pumpkin fruit and stem instance segmentation

The agricultural environment is a typical unstructured environment with complex background. Due to 'free growth' and overlapped fruits, stems, branches, and leaves, fine instance segmentation in fruit harvesting environment becomes a challenging work. In this paper, we proposed a pumpkin fruit and stem instance segmentation framework as illustrated in Figure 5. The main feature of this framework is introducing a transformer network to replace the commonly used convolutional

neural network (CNN). The transformer network helps effectively extract image features, improve instance segmentation accuracy, and reduce model computational complexity. In addition, to deal with the overlapping phenomenon that often occurs in the harvesting environment, we add a hard point selection module to the mask branch. Coarse features are concatenated with fine features from the output of the feature pyramid network (FPN) to classify those hard points and then generate the final fine mask.

Compared with the literature in the previous review, our framework achieves the end-to-end fruit and stem instance segmentation. The specific implementation is as follows. First, we introduce the Swin Transformer (Liu et al., 2021b) to the task of pumpkin fruit and stem instance segmentation and replace CNN (such as ResNet) to extract features. The feature extraction structure of this transformer network combined with a feature pyramid network (FPN) (Lin et al. (2017)). Hierarchical transformer and FPN are applied to generate a pyramid of feature maps with different sizes of a fixed number of channels (set to 256). Specially, we use four levels of feature maps denoted as $\{P2,P3,P4, P5\}$. $P2$, $P3$, $P4$ and $P5$ are generated by four transformer feature maps $T2$, $T3$, $T4$ and $T5$, convolving with 1×1 kernel *via* top-down connection mechanism. As a result, $P2$, $P3$, $P4$ and $P5$ have strides 4, 8, 16 and 32 respectively. Then a region proposal network (RPN) (Ren et al., 2015) is deployed to generate the feature map with anchors. *Via* RoIAlign (He et al., 2017) operation, fixed-size feature maps can be obtained. After fully connected (FC) layers, prediction results of the bounding box and classification are output. In a general way, fixed-size feature maps can generate mask predictions after several convolution operations. However, since the fixed size of the feature map is generally 7*7, it is tough to generate an
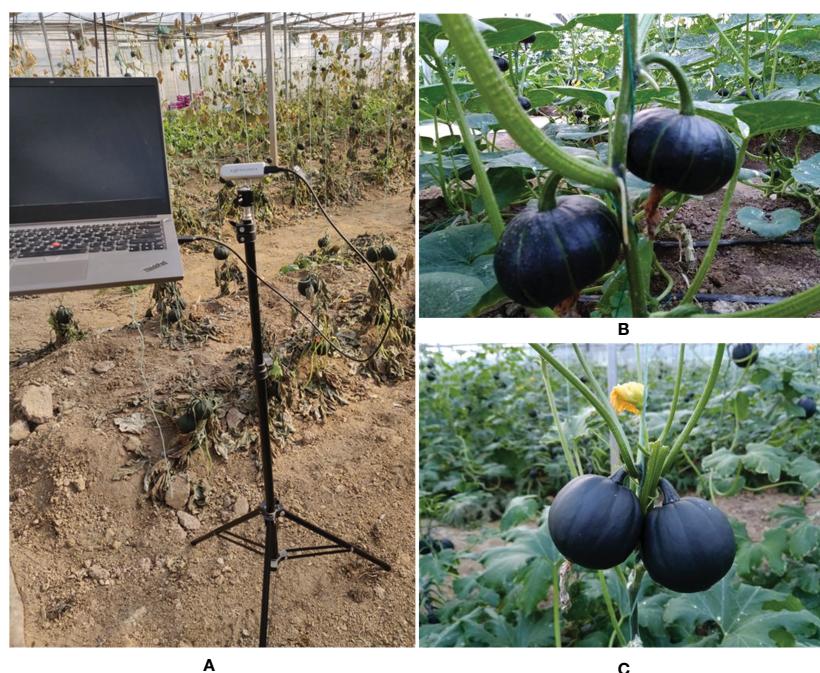

FIGURE 2
Image collection environment and pumpkin images. **(A)** Image capture scene. **(B)** Original Bebe pumpkin image. **(C)** Original Hazel pumpkin image.

FIGURE 3
Illustration of the image annotation process. **(A)** Original image. **(B)** Polygonal annotation and extraction of the pumpkin fruit mask. **(C)** Polygonal annotation and extraction of the pumpkin stem mask. **(D)** Annotated images, red polygons are pumpkin fruits, green polygons are pumpkin stems, yellow boxes are pumpkins.

accurate mask, especially at the fruit boundary. Therefore, we select the hard points in the edge area and combine the coarse features from the fixed-size feature map and the fine features from high-resolution feature map output by FPN to generate more refined point-wise label predictions. Details of transformer network and mask branch will be introduced in subsections.

2.2.1 Transformer network

Transformer has a great impact on the field of natural language processing (NLP) before. The proposal of vision transformer (ViT) (Dosovitskiy et al., 2020) breaks the gap between NLP and vision, and replaces the convolutional neural network with a pure transformation module to perform image classification tasks. Liu et al. (2021b) proposed a new visual transformer, called Swin Transformer, whose multi-scale and computationally inexpensive properties make it compatible with a wide range of vision tasks (image classification, object detection, semantic segmentation, etc.). An overview of the transformer architecture and transformer blocks we

applied are presented in Figure 6. It first splits an input RGB image into non-overlapping patches (raw-valued features) by a patch partition operation. Then a linear embedding layer is applied to the raw-valued features to project them to an arbitrary dimension (set to 96). Several transformer blocks are applied to these patch tokens. To produce a hierarchical representation, the number of tokens is reduced by patch merging layers as the network gets deeper. Specific implementations are demonstrated in (Liu et al., 2021b).

2.2.2 Mask branch

In the instance segmentation task of agricultural environment, due to the large-scale overlapping problem, fine segmentation of the target edges and overlapping edges is challenging. Research in (Li et al. 2017) shows that in the segmentation task, most of the hard pixels (about 70%) are at the edge of the object. Point rendering method (Kirillov et al., 2020) we applied is devoted to segmenting these blurry pixels finely. Figure 7 depicts the main idea of point rendering. Point rendering includes three steps:

TABLE 1 Distribution of the dataset.

	Images	Fruit instances	Stem instances	Pumpkin bounding boxes
Bebe	354	608	552	608
Hazel	325	676	516	676
Total	679	1284	1068	1284
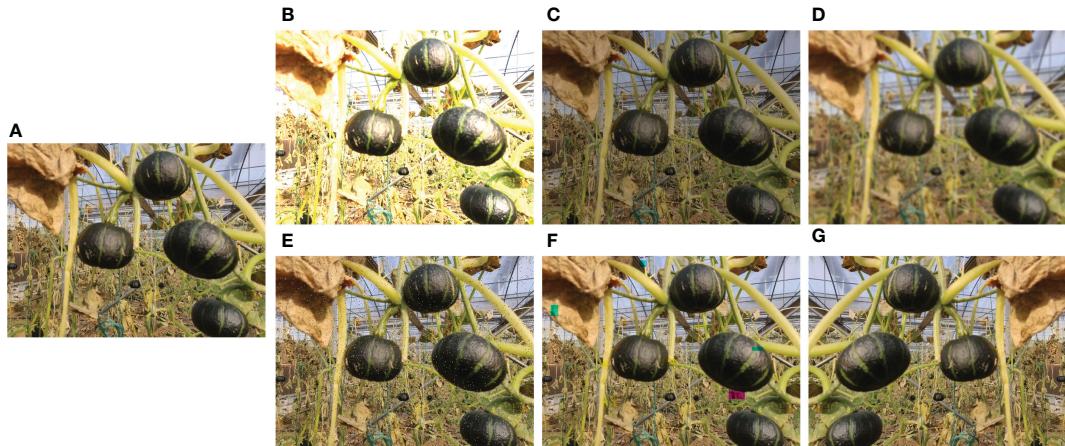
FIGURE 4
Data augmentation. **(A)** Original image; **(B)** Brightness enhancement; **(C)** Brightness reduction; **(D)** Gaussian blur; **(E)** Noise; **(F)** Cutout; **(G)** horizontal flip.

2.2.2.1 Candidate point selection

First, *via* upsampling, the low-resolution segmentation map is converted to high-resolution, and N hard points with low confidence are filtrated in the high-resolution segmentation map. Most of these points are concentrated near the edge. This process iterates step by step to obtain a segmentation map of the desired resolution. In the implementation of this paper, the N value we choose is 28*28.

2.2.2.2 Point feature extraction

Coarse and fine features for each candidate point are extracted. The coarse features are extracted from the low-resolution segmentation map, and the fine features are taken from the P2 layer of the FPN. The two sets of features are concatenated to obtain the feature expression of the candidate points.

2.2.2.3 Point prediction

After obtaining the features of the candidate points, through a set of multi-layer perceptions (MLP), the final segmentation prediction results of the candidate points are obtained. More implementation details can be seen in Kirillov et al. (2020).

2.2.3 Training and inference
2.2.3.1 Training

In our implementation, we apply a multi-scale training mechanism (He et al. 2015). To address the issue of varying image sizes in training. In each epoch, a scale is randomly selected for training.

In the proposed pumpkin fruit and stem instance segmentation network, we define the training loss function as Equation (1):

$$L = L_{classification} + \lambda L_{box} + \gamma L_{mask} \tag{1}$$
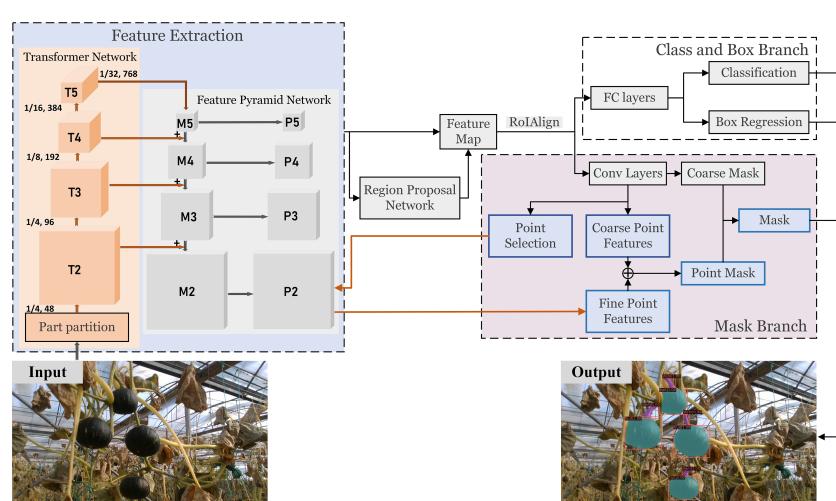
FIGURE 5
Pumpkin fruit and stem instance segmentation framework.
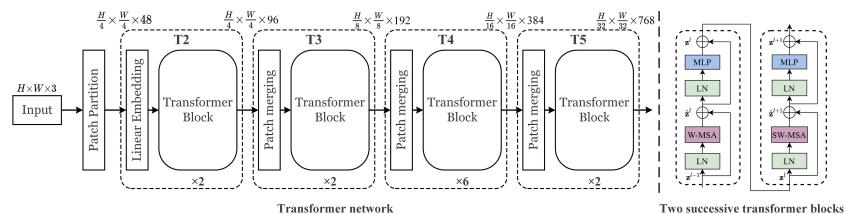
FIGURE 6
The architecture of transformer network.

where $L_{classification}$ is the loss for fruit or stem classification, L_{box} is the loss for the bounding box coordinates prediction, and L_{mask} is the loss for mask prediction.

In our implementation, we apply cross entropy loss to calculate $L_{classification}$ and L_{mask}, L1 loss to calculate L_{box}. We set λ to 1 and γ to 2 because mask is more difficult to train and is more important in our implementation.

2.2.3.2 Inference

The inference of the pumpkin fruit and stem instance segmentation network is a straightforward process. We forward input images through the transformer backbone and FPN. We select the points from the 224×224 resolution feature map refined by the coarse 7×7 prediction in 5 steps. We select the $N=28^2$ most uncertain points based on the absolute difference between the predictions and 0.5.

2.3 Cutting and grasping point estimation

The proposed pumpkin keypoint detection framework is illustrated in Figure 8. Firstly, fruit and stem masks are generated *via* instance segmentation method as shown in Figure 8B. After obtaining the instance segmentation result, the fruit instances and the stem instances can be separated as depicted in Figure 8C. Among these instances, there are corresponding relationships among the fruits and the stems, and only

one-to-one fruit and stem can be labeled as the pumpkin picking target. Then, we apply a geometric model to determine the cutting and grasping points. Finally, by modeling the robot and its coordinate systems, calibrating the camera parameters, the target pixel in 2D image can be transformed a position in 3D space. In practical operations, Birrell et al. (2020); Wang et al. (2022) and Kang et al. (2020) proposed approaches to tackle the coordinate transformation problem. Two pivotal steps of the keypoint estimation algorithm are fruit and stem correspondence determination and keypoint determination.

2.3.1 Fruit and stem correspondence determination

To determine the corresponding relationships among the fruit and stem instances, we take advantage of the apriori knowledge. Three requirements should be satisfied. 1) The masks of fruit and stem are adjacent. 2) Under the force of gravity, the center point of the stem is above the center point of the fruit. 3) One fruit corresponds to at most one stem. Algorithm 1 shows the matching process.

2.3.2 Keypoint determination

After obtaining the mask of a whole pumpkin includes a fruit and a corresponding stem by the proposed correspondence determination algorithm, a geometric model is employed to
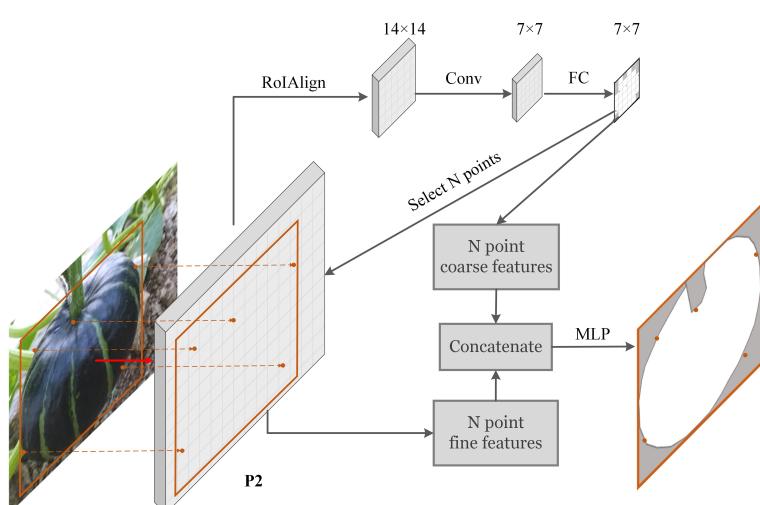
FIGURE 7
Scheme of point rendering mask head.
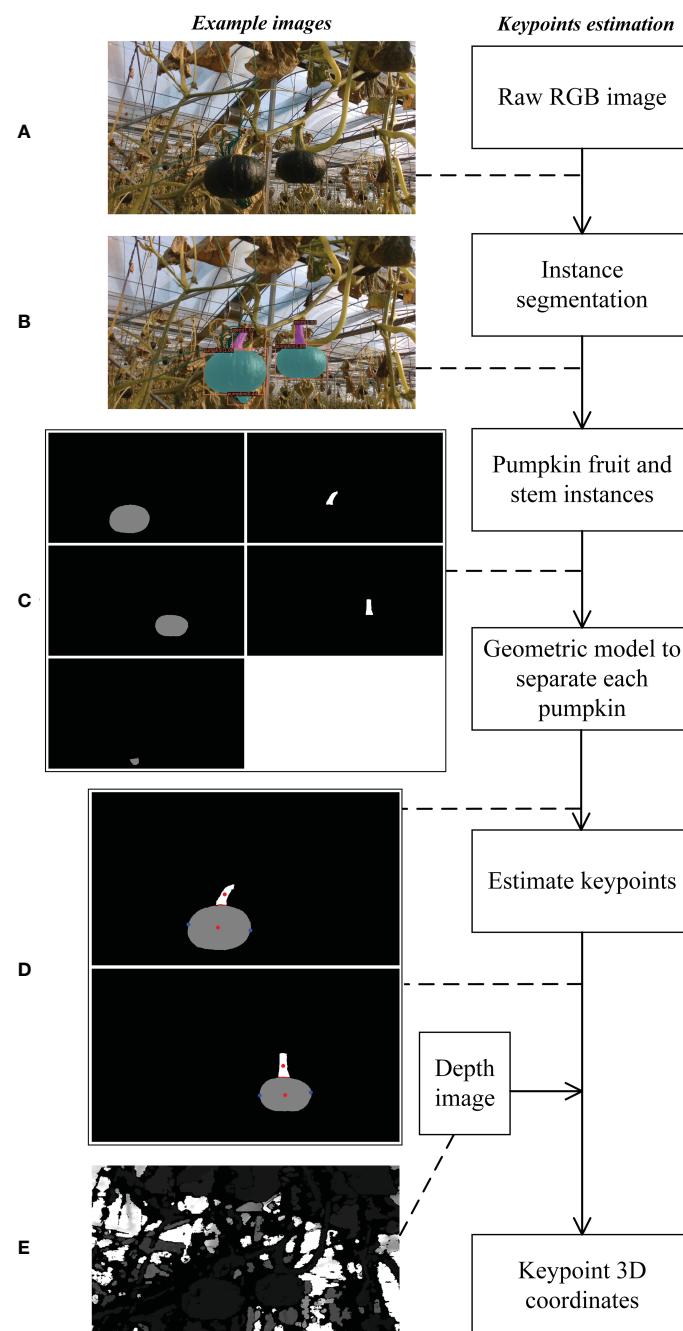
FIGURE 8
Block diagram of cutting and grasping point estimation method along with example images. **(A)** Pumpkin RGB image. **(B)** Pumpkin fruit and stem instance segmentation result. **(C)** Visualization of pumpkin fruit and stem instances. The left column instances are fruits, and the right column instances are stems. **(D)** The separate pumpkins with keypoints. The red points between the stem and the fruit are adjacent points. The red dot in the stem is the cutting point, and the blue points in the fruit are the grasping points. **(E)** Pumpkin depth image. Best viewed zoom in.

estimate the exact location of the grasping points and cutting point.

Denote fruit mask as $F: \{(x_i^F, y_i^F)\}_{i=1}^{N_F}$, stem mask as $S: \{(x_i^S, y_i^S)\}_{i=1}^{N_S}$, where N_F and N_S represents number of fruit pixels and stem pixels respectively. As illustrated in Figure 9, first, the center of mass of the 2D fruit and stem is calculated as Equation (2), labeled as

$\{C_F : (x_{cf}, y_{cf})\}$ and $\{C_S : (x_{cs}, y_{cs})\}$ respectively.

$$x_{c_f} = \frac{\sum_{i=1}^{N_F} x_i^F}{N_F}, \ y_{c_f} = \frac{\sum_{i=1}^{N_F} y_i^F}{N_F}, \ x_{c_s} = \frac{\sum_{i=1}^{N_S} x_i^S}{N_S}, \ y_{c_s} = \frac{\sum_{i=1}^{N_S} y_i^S}{N_S} \quad (2)$$

A straight line denoted as l_p passes through C_F and C_S. Considering two conditions:

```
Input:
Fruit^i ∈ ℕ^{H×W}: The i-th fruit mask in the image;
Stem^j ∈ ℕ^{H×W}: The j-th stem mask in the image;
M: Number of fruits detected in the image;
N: Number of stems detected in the image;
Output:
K matched pairs, each pair has a fruit mask and a
stem mask,
K ≤ min{M,N}
1: for i=0 to M do
2: Calculate center point of Fruit^i: C_{F_i} = (x_i, y_i)
3:   for j=0 to N do
4:   Calculate center point of Stem^j: C_{S_j} = (x_j, y_j)
5:   if x_j > x_i (To ensure the center point of stem
     is above the center point of fruit) then
6:     count adjacent mask point: num(dis < dis_thr)
7:     if num(dis < dis_thr) > num_thr (To ensure the
       masks of fruit and stem are adjacent) then
8:       Fruit^i and Stem^j is a matching candidate
9:     end if
10:  end if
11:  for i=0 to M do
12:    if There is one or more than one match
       candidates with stem then
13:      Calculate the degree of pumpkin matching
         candidate: D = arctan (|x_i - x_j|)/(|y_i - y_j|)
14:      The matching candidate with the minimum D
         value is determined as the match pair. (To
         ensure one fruit corresponds to at most one
         stem)
15:    end if
16:  end for
17: end for
18: end for
```

ALGORITHM 1
Matching pumpkin fruit masks with their corresponding stem masks.

Case 1: l_p is a vertical line (slope of l_p is ∞).

Denote l_p as $x=c$. Index of grasping points G_1 and G_2 from fruit mask F can be calculated as Equation (3):

$$\begin{cases} G_1 = \arg \max_i \ \left|x_i^F - c\right|, \quad x \le c \\ G_2 = \arg \max_i \ \left|x_i^F - c\right|, \quad x > c \end{cases} \qquad (3)$$

Case 2: l_p is not a vertical line (slope of l_p is not ∞).

Assume line equation as $l_p : y = kx + b$. Denote D_i as the distance from ith point in F to l_p. G_1 and G_2 can be calculated as Equation (4):

$$\begin{cases} G_1 = \arg \max_i \ D_i, \quad y_i^F \le kx_i^F + b \\ G_2 = \arg \max_i \ D_i, \quad y_i^F > kx_i^F + b \end{cases} \qquad (4)$$

Finally, $keypoint_{G_1} : x_{G_1}^F, y_{G_1}^F$ and $keypoint_{G_2} : x_{G_2}^F, y_{G_2}^F$ are determined as two grasping points that distribute in two sides of

l_p. Cutting point $keypoint_C$ is the center point of stem mask, that is $\{C_S : (x_{c_s}, y_{c_s})\}$.

3 Results and discussion

3.1 Evaluation metrics

3.1.1 Average precision

According to whether the true sample and the predicted result match, the prediction results can be divided into four types: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). Precision and recall are defined as follows:

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

The average precision metric is used to measure the quality of the detections and the segmentations obtained by the models. Average precision computes the average precision value for recall values over 0 to 1. Specifically, mean average precision (mAP) is defined as the primary metric. As in (Lin et al. (2014)), mAP is calculated using 10 intersection over union (IoU) thresholds from 0.50 to 0.95. The IoU measures the overlap between two boundaries or masks and measures how much the box boundary or mask predicted by the algorithm overlaps with the ground truth (the real object boundary or real object mask).

3.1.2 Model complexity and inference speed

The model complexity usually relates to parameter number and calculation amount, two metrics that describe how many parameters the model defines and how many floating point
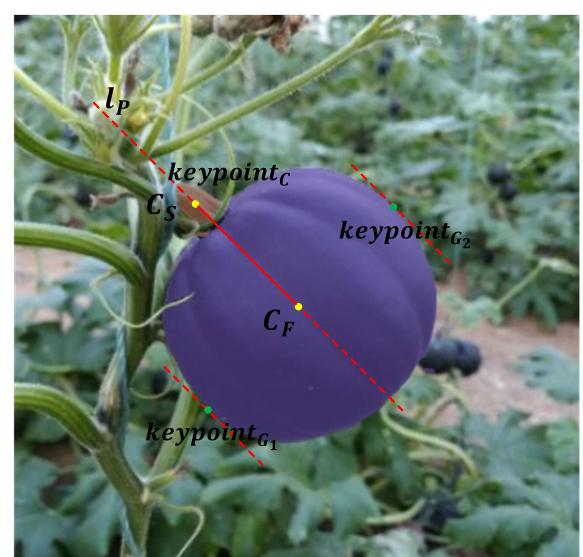

FIGURE 9
Estimation of pumpkin grasping and cutting points.

operations(FLOPs) are required when running the model. 1GFLOPs = 10^9FLOPs. The metric to define the model inference speed is the average number of frames per second (FPS). Model complexity and FPS are vital indicators to evaluate the performance of the model.

3.2 Instance segmentation result

3.2.1 Experiment setup

In this paper, the training and evaluation of the proposed network are conducted on a server, which consists of an Intel i9-10900X CPU with 20 cores, 32G RAM, and an RTX 3090 GPU with 24G memory. The network implementation was carried out using Pytorch 1.7.0.

3.2.2 Performance comparison with state-of-the-art methods

We performed a series of experiments to compare our method with the state-of-the-art methods, namely YOLACT (Bolya et al., (2019), QueryInst (Fang et al., 2021), Mask R-CNN (He et al., (2017) and Cascade Mask R-CNN (Cai and Vasconcelos, 2019). All algorithms are trained for 100 epochs, and when every training epoch ends, the mAP values of mask segmentation and box detection are calculated as shown in Figures 10, 11. The detection mAP of our proposed method outperforms these state-of-the-art methods, and the segmentation mAP is significantly superior to the existing methods. Fortunately, in this application, segmentation precision is more important than detection precision.

The evaluation results are listed in Table 2. Our architecture achieves a high instance segmentation accuracy of 0.708 mask mAP and 0.720 box mAP, which brings 4.9% and 2.5% gains over the second-best results. From the parameter comparison, except QueryInst (the model is too large) and YOLACT (the accuracy is not satisfactory), the margin among parameter numbers of Mask R-CNN, Cascade Mask R-CNN and the method we proposed is narrow. It's worth noting that although the parameter size of our method is larger than Mask R-CNN (59.27M Vs 43.76M), the computational complexity is lower than Mask R-CNN (213.01 GFLOPs Vs 258.19 GFLOPs). Our method achieves 13.5 FPS on a single RTX 3090 GPU, which can meet the requirements of agricultural applications.

3.2.3 Visualization result analysis

To highlight the superiority of the proposed architecture more intuitively, the visual analysis of the outstanding networks and our network is conducted. As can be seen in Figure 12, all methods can detect the majority of pumpkin instances, whereas our method achieves higher confidence. As is shown in the third column, YOLACT and QueryInst fail to detect the pumpkin in red circle covered by the leave, while Mask R-CNN, Cascade Mask R-CNN and our method detect the pumpkin with the confidence of 0.38, 0.97 and 1.0, respectively. It is obvious that our method generates finer masks compared with other methods. To emphasize the contribution of point rendering mask branch, we compared the visualization results of our method and our

FIGURE 10
Mask segmentation mAP of the model.
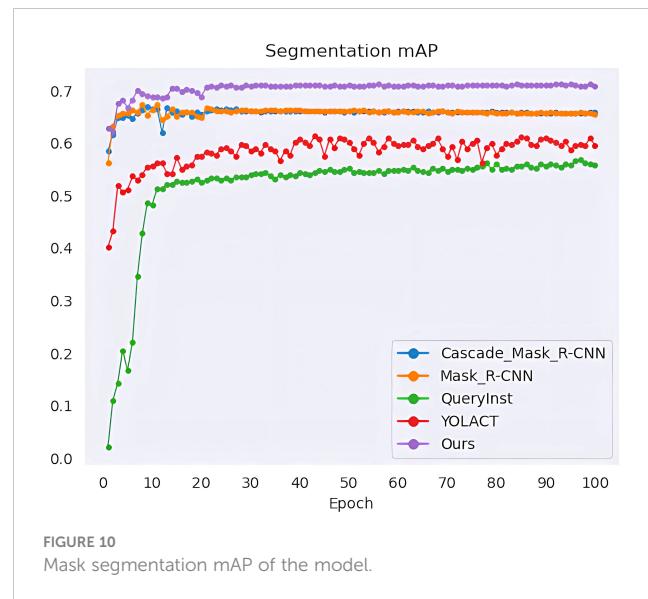
method without point rendering as shown in the last two rows, where can be seen that the finer masks benefit more from the point rendering mechanism.

3.2.4 Ablation study of improved models

In order to prove the effectiveness of the improved modules in the proposed pumpkin fruit and stem instance segmentation method, the ablation study on different modules is performed in this section. The comparisons are conducted on seven cases, as shown in Table 3. As can be seen from the table, replacing the original CNN module with the transformer network and the mechanism of multi-scale training have greatly improved the results, and the mask mAP and box mAP have increased by 2.5%, 2.4% and 1.7%, 3.3% respectively. Although the improvement of replacing the original mask branch with the point rendering mask branch takes no remarkable superiority in mAP results, it only increases by 0.6% in mask mAP, and the box mAP has a slight

FIGURE 11
Box detection mAP of the model.
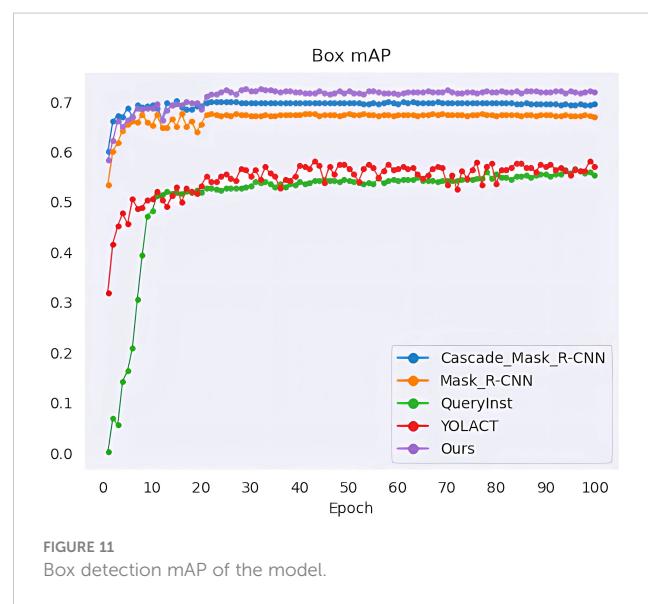
TABLE 2 Performance comparison with state-of-the-art methods.

	mask mAP	box mAP	#param.	GFLOPs	FPS
YOLACT (Bolya et al. (2019))	0.596	0.572	**34.74M**	**186.57**	**21.3**
QueryInst (Fang et al. (2021))	0.559	0.554	172.23M	464.29	6.2
Mask R-CNN (He et al. (2017))	0.656	0.669	43.76M	258.19	16.4
Cascade Mask R-CNN (Cai and Vasconcelos (2019))	0.659	0.695	76.8M	389.03	13.6
Ours	**0.708**	**0.720**	59.27M	213.01	13.5

The best performances of each metrics are in bold format.

increase of 0.3%, but from the visualization results, point rendering mask branch greatly optimizes the boundary masks, which cannot be ignored. Finally, the architecture with transformer network, point rendering mask branch, and the multi-scale training network improves 5.3% mask mAP and 5.1% box mAP over the Mask R-CNN Baseline network. The inference speed decreased from 16.4 FPS to 13.5 FPS, but this is acceptable.

3.3 Keypoint estimation results

3.3.1 Pumpkin fruit and stem correspondence determination result

Figure 13 shows some example results of fruit and stem matching algorithm. It can be seen that in most conditions, including one image with single or multiple pumpkins, existing
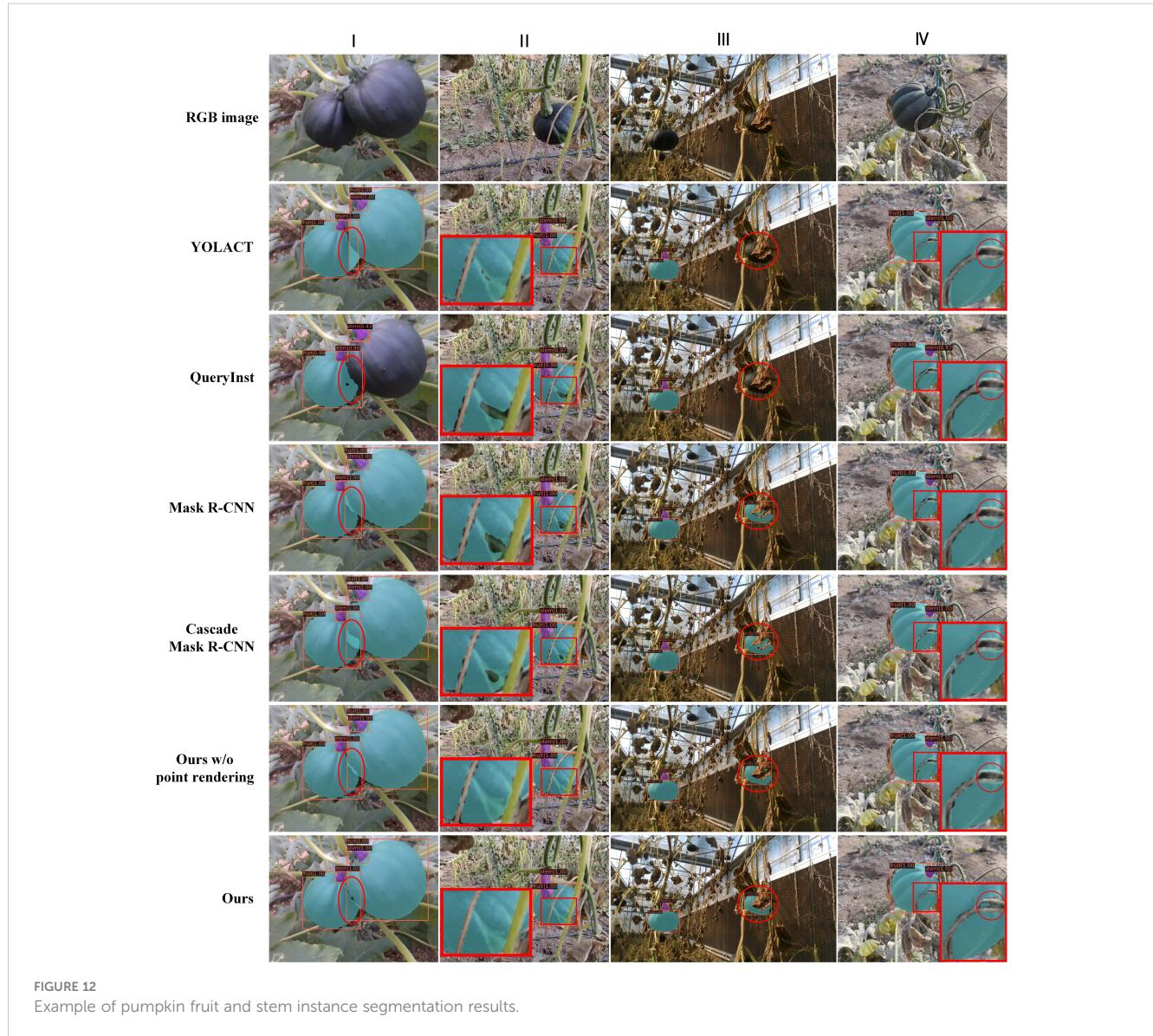
FIGURE 12
Example of pumpkin fruit and stem instance segmentation results.

TABLE 3   Ablation study on the pumpkin fruit and stem instance segmentation method.

Model	transformer network	multi-scale training	point rendering mask	mask mAP	box mAP	FPS
Baseline model				0.656	0.669	16.4
Model-A	√			0.681	0.686	15.4
Model-B		√		0.680	0.702	**16.5**
Model-C			√	0.665	0.672	14.3
Model-D	√	√		0.705	0.718	15.1
Model-E		√	√	0.701	0.709	13.7
Model-F (Ours)	√	√	√	**0.708**	**0.720**	13.5

The best performances of each metrics are in bold format.

fruit, leave, or branch overlaps, our algorithm can match the fruits and stems successfully. To analyze the results accurately, we count all the matched pumpkin instances in the test images, the number of TP is 215, FP is 4, and TN is 2. The precision and recall reach 98.2% and 99.1% respectively. Some negative matched examples are listed in Figure 14. The reason for the faults is that in the instance segmentation step, missing and erroneous detections happen sometimes. The pumpkin is too small or interference of branches may cause false detection.

3.3.2 Keypoint determination result

Figure 15 presents the visualized results of grasping and cutting keypoint detection. In the figure, red points are the cutting points, blue points are the grasping points. Yellow lines linking the cutting points and grasping points signify that the 3 points attach to one pumpkin instance. Specially, the occlusion problem is usually not negligible in fruit picking task. One of the advantages of this approach is that if a pumpkin is occluded seriously, for instance,

FIGURE 13
Pumpkin fruit and stem matching result.
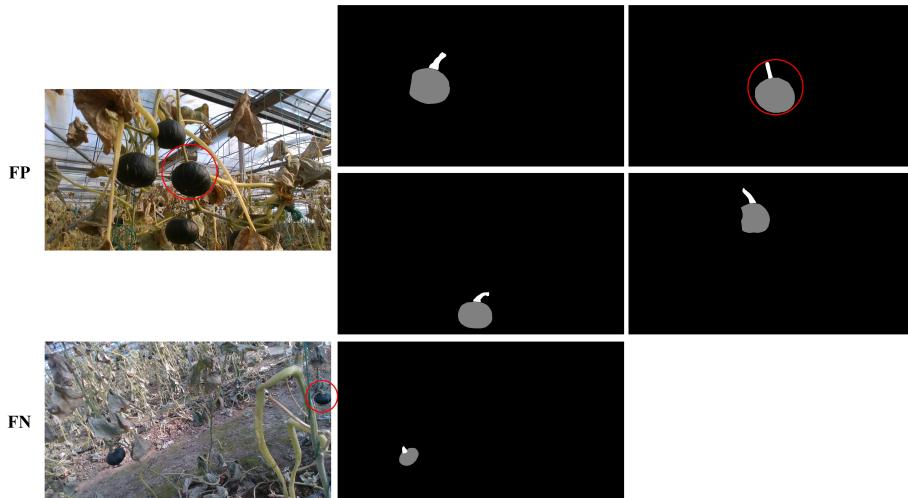
FIGURE 14

Wrong examples of pumpkin fruit and stem matching result.

if we can only see the fruit part or the stem part in the image, our algorithm can filter this pumpkin autonomous as shown in the first image from the second row in Figure 15. If the pumpkin is only occluded part of the fruit or stem by leaves, branched or other fruits, our algorithm also determines the grasping point and cutting point reasonably as shown in the right three columns from Figure 15. The results show that our algorithm is promising for the pumpkin picking task.

4 Conclusion

In this paper, we presented a pumpkin autonomous picking framework with keypoint detection and instance segmentation method. A transformer network is utilized as the architecture backbone to replace CNN, which helps achieve a higher detection and segmentation precision. To tackle the overlapping problem, point rendering is applied so that finer masks can be acquired. Sufficient experimental results indicate that our method significantly outperforms several state-of-the-art instance segmentation methods. In addition, a novel keypoint detection algorithm is proposed to model the relationships among the fruit and stem instances as well as estimate grasping and cutting keypoints. The effectiveness and applicability of the proposed method are verified through plenty experiments on pumpkin image dataset we created. In this work, we applied traditional geometric method to model the fruit-stem relationships and estimating the keypoints. Our future work will expand into learning-based method to detect the fruit-stem pairs and directly generate the keypoints using deep neural networks.

FIGURE 15

Visualized results of grasping and cutting keypoint detection.

Data availability statement

Author contributions

JY, YL, DZ, and TX participated in the conception and design of this research and revised the manuscript. JY carried out the experiments, organized the database, performed data analysis, and wrote the manuscript. YL, DZ, and TX advised on the design of the model and analyzed to find the best method for this work. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

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

Publisher's note

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

References

Bhatti, U. A., Tang, H., Wu, G., Marjan, S., and Hussain, A.. (2023). Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence. *Int. J. Intelligent Syst.* 2023, 1–28. doi: 10.1155/2023/8342104

Birrell, S., Hughes, J., Cai, J. Y., and Iida, F. (2020). A field-tested robotic harvesting system for iceberg lettuce. *J. Field Robotics* 37, 225–245. doi: 10.1002/rob.21888

Bolya, D., Zhou, C., Xiao, F., and Lee, Y. J. (2019). "Yolact: Real-time instance segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*. 9157–9166. doi: 10.1109/ICCV.2019.00925

Bresilla, K., Perulli, G. D., Boini, A., Morandi, B., Corelli Grappadelli, L., and Manfrini, L. (2019). Single-shot convolution neural networks for real-time fruit detection within the tree. *Front. Plant Sci.* 10, 611. doi: 10.3389/fpls.2019.00611

Cai, Z., and Vasconcelos, N. (2019). Cascade R-CNN: high quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 1483–1498. doi: 10.1109/TPAMI.2019.2956516

Chen, Y.-T., and Chen, S.-F. (2020). Localizing plucking points of tea leaves using deep convolutional neural networks. *Comput. Electron. Agric.* 171, 105298. doi: 10.1016/j.compag.2020.105298

Chen, T., Zhang, R., Zhu, L., Zhang, S., and Li, X. (2021). A method of fast segmentation for banana stalk exploited lightweight multi-feature fusion deep neural network. *Machines* 9, 66. doi: 10.3390/machines9030066

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* 8, 2955–2969.

Fang, Y., Yang, S., Wang, X., Li, Y., Fang, C., Shan, Y., et al. (2021). "Instances as queries," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6910–6919. doi: 10.1109/ICCV48922.2021.00683

Fu, L., Majeed, Y., Zhang, X., Karkee, M., and Zhang, Q. (2020). 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

Ganesh, P., Volle, K., Burks, T., and Mehta, S. (2019). Deep orange: Mask R-CNN based orange detection and segmentation. *IFAC-PapersOnLine* 52, 70–75. doi: 10.1016/j.ifacol.2019.12.499

Gonzalez, S., Arellano, C., and Tapia, J. E. (2019). Deepblueberry: Quantification of blueberries in the wild using instance segmentation. *IEEE Access* 7, 105776–105788. doi: 10.1109/ACCESS.2019.2933062

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," in *Proceedings of the IEEE international conference on computer vision*. 2961–2969. doi: 10.1109/ICCV.2017.322

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

Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., and Zheng, Y. (2020). Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Comput. Electron. Agric.* 172, 105380. doi: 10.1016/j.compag.2020.105380

Kalampokas, T., Vrochidou, E., Papakostas, G. A., Pachidis, T., and Kaburlasos, V. G. (2021). Grape stem detection using regression convolutional neural networks. *Comput. Electron. Agric.* 186, 106220. doi: 10.1016/j.compag.2021.106220

Kang, H., Zhou, H., Wang, X., and Chen, C. (2020). Real-time fruit recognition and grasping estimation for robotic apple harvesting. *Sensors* 20, 5670. doi: 10.3390/s20195670

Kirillov, A., Wu, Y., He, K., and Girshick, R. (2020). "Pointrend: Image segmentation as rendering," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9799–9808. doi: 10.1109/CVPR42600.2020.00982

Li, X., Liu, Z., Luo, P., Change Loy, C., and Tang, X. (2017). "Not all pixels are equal: Difficulty-aware semantic segmentation *via* deep layer cascade," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3193–3202. doi: 10.1109/CVPR.2017.684

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2117–2125. doi: 10.1109/CVPR.2017.106

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft Coco: Common objects in context," in *European conference on computer vision*. 740–755 (Springer).

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021b). "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10012–10022. doi: 10.1109/ICCV48922.2021.00986

Liu, G., Mao, S., and Kim, J. H. (2019a). A mature-tomato detection algorithm using machine learning and color analysis. *Sensors* 19, 2023. doi: 10.3390/s19092023

Liu, Y., Zhang, Z., Liu, X., Wang, L., and Xia, X. (2021a). Efficient image segmentation based on deep learning for mineral image classification. *Advanced Powder Technol.* 32, 3885–3903. doi: 10.1016/j.apt.2021.08.038

Liu, X., Zhao, D., Jia, W., Ji, W., Ruan, C., and Sun, Y. (2019b). Cucumber fruits detection in greenhouses based on instance segmentation. *IEEE Access* 7, 139635–139642. doi: 10.1109/ACCESS.2019.2942144

Luo, L., Tang, Y., Lu, Q., Chen, X., Zhang, P., and Zou, X. (2018). A vision methodology for harvesting robot to detect cutting points on peduncles of double overlapping grape clusters in a vineyard. *Comput. industry* 99, 130–139. doi: 10.1016/j.compind.2018.03.017

Midtiby, H. S., and Pastucha, E. (2022). Pumpkin yield estimation using images from a UAV. *Agronomy* 12, 964. doi: 10.3390/agronomy12040964

Pan, C., Yang, K., Erhunmwunsee, F., Li, Y.-X., Liu, M., Pan, S., et al. (2023). Inhibitory effect of cinnamaldehyde on fusarium solani and its application in postharvest preservation of sweet potato. *Food Chem.* 408, 135213. doi: 10.1016/j.foodchem.2022.135213

Perez-Borrero, I., Marin-Santos, D., Gegundez-Arias, M. E., and Cortes-Ancos, E. (2020). A fast and accurate deep learning method for strawberry instance segmentation. *Comput. Electron. Agric.* 178, 105736. doi: 10.1016/j.compag.2020.105736

Perez-Borrero, I., Marin-Santos, D., Vasallo-Vazquez, M. J., and Gegundez-Arias, M. E. (2021). A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. *Neural Computing Appl.* 33, 15059–15071. doi: 10.1007/s00521-021-06131-2

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, 1–9.

Rong, J., Dai, G., and Wang, P. (2021). A peduncle detection method of tomato for autonomous harvesting. *Complex Intelligent Syst.* 1–15. doi: 10.1007/s40747-021-00522-7

Sa, I., Lehnert, C., English, A., McCool, C., Dayoub, F., Upcroft, B., et al. (2017). Peduncle detection of sweet pepper for autonomous crop harvesting combined color and 3-d information. *IEEE Robotics Automation Lett.* 2, 765–772. doi: 10.1109/LRA.2017.2651952

Sun, Q., Chai, X., Zeng, Z., Zhou, G., and Sun, T. (2021). Multi-level feature fusion for fruit bearing branch keypoint detection. *Comput. Electron. Agric.* 191, 106479. doi: 10.1016/j.compag.2021.106479

Sun, J., Jia, Q., Li, Y., Zhang, T., Chen, J., Ren, Y., et al. (2022). Effects of arbuscular mycorrhizal fungi and biochar on growth, nutrient absorption, and physiological properties of maize (zea mays l.). *J. Fungi* 8, 1275. doi: 10.3390/jof8121275

Sun, S., Jiang, M., He, D., Long, Y., and Song, H. (2019). Recognition of green apples in an orchard environment by combining the GrabCut model and ncut algorithm. *Biosyst. Eng.* 187, 201–213. doi: 10.1016/j.biosystemseng.2019.09.006

Wan, H., Fan, Z., Yu, X., Kang, M., Wang, P., and Zeng, X. (2022). A real-time branch detection and reconstruction mechanism for harvesting robot *via* convolutional neural network and image segmentation. *Comput. Electron. Agric.* 192, 106609. doi: 10.1016/j.compag.2021.106609

Wang, X., Kang, H., Zhou, H., Au, W., and Chen, C. (2022). Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. *Comput. Electron. Agric.* 193, 106716. doi: 10.1016/j.compag.2022.106716

Wittstruck, L., Kühling, I., Trautz, D., Kohlbrecher, M., and Jarmer, T. (2020). UAV-based RGB imagery for hokkaido pumpkin (cucurbita max.) detection and yield estimation. *Sensors* 21, 118. doi: 10.3390/s21010118

Yang, K., Geng, Q., Luo, Y., Xie, R., Sun, T., Wang, Z., et al. (2022). Dysfunction of fada-camp signalling decreases aspergillus flavus resistance to antimicrobial natural preservative perillaldehyde and afb1 biosynthesis. *Environ. Microbiol.* 24, 1590–1607. doi: 10.1111/1462-2920.15940

Yoshida, T., Fukao, T., and Hasegawa, T. (2018). Fast detection of tomato peduncle using point cloud with a harvesting robot. *J. Robotics Mechatronics* 30, 180–186. doi: 10.20965/jrm.2018.p0180

Yuan, T., Lv, L., Zhang, F., Fu, J., Gao, J., Zhang, J., et al. (2020). Robust cherry tomatoes detection algorithm in greenhouse scene based on SSD. *Agriculture* 10, 160. doi: 10.3390/agriculture10050160

Different leafy vegetable cropping systems regulate growth, photosynthesis, and PSII functioning in mono-cropped eggplant by altering chemical properties and upregulating the antioxidant system

Muhammad Imran Ghani[1,2,3], Ahmad Ali[3],
Muhammad Jawaad Atif[3,4], Muhammad Ali[3],
Mohammad Abass Ahanger[5], Xiaoyulong Chen[1,2,6]*
and Zhihui Cheng[3]*

[1]College of Agriculture, Guizhou University, Guiyang, China, [2]Key Laboratory of Karst Georesources and Environment, Ministry of Education, College of Resources and Environmental Engineering, Guizhou University, Guiyang, China, [3]College of Horticulture, Northwest A&F University, Yangling, China, [4]Horticultural Research Institute, National Agricultural Research Centre, Islamabad, Pakistan, [5]College of Life Sciences, Northwest A&F University, Yangling, China, [6]College of Ecology and Environment, Tibet University, Lhasa, Tibet, China

Continuous cropping of eggplant threatened regional ecological sustainability by facilitating replanting problems under mono-cropping conditions. Therefore, alternative agronomic and management practices are required to improve crop productivity at low environmental cost for the development of sustainable agricultural systems in different regions. This study examined changes in soil chemical properties, eggplant photosynthesis, and antioxidant functioning in five different vegetable cropping systems over a 2-year period., 2017 and 2018. The results showed that welsh onion-eggplant (WOE), celery-eggplant (CE), non-heading Chinese cabbage-eggplant (NCCE), and leafy lettuce-eggplant (LLE) rotation systems significantly impacted growth, biomass accumulation, and yield than fallow-eggplant (FE). In addition, various leafy vegetable cropping systems, WOE, CE, NCCE, and LLT induced significant increases in soil organic matter (SOM), available nutrients (N, P, and K), and eggplant growth by affecting the photosynthesis and related gas exchange parameters with much evident effect due to CE and NCCE. Moreover, eggplant raised with different leafy vegetable rotation systems showed higher activity of antioxidant enzymes, resulting in lower accumulation of hydrogen peroxide and hence reduced oxidative damage to membranes. In addition, fresh and dry plant biomass was significantly increased due to crop rotation with leafy vegetables. Therefore, we concluded that leafy vegetable crop rotation is a beneficial management practice to improve the growth and yield of eggplant.

1 Introduction

The consistent increase in the global population has increased the demand for food and cash crops. Due to limited agricultural land worldwide and the decreasing number of new crop areas, monoculture is a common model for large-scale, intensive agricultural production, especially in the horticultural industry (Tan et al., 2021; Bhatti et al., 2022). Even with a good field management regime, the crop may still experience growth and yield reduction and promote disease incidence (Zeng et al., 2020). Mono-cropping refers to the practice of growing the same type of crop every year for a long period of time (Scarascia-Mugnozza et al., 2011; Xiao et al., 2012). The overuse of synthetic fertilizers and agrochemicals is inevitable in the mono-cropping system, which leads to mono-cropping obstacles (Ghani et al., 2019b; Ali et al., 2021; Ghani et al., 2022b). The mono-cropping obstacles are attributed to soil salinization, acidification, nutrient imbalance, and autotoxicity (Lyu et al., 2020; Zeng et al., 2020).

Eggplant (*Solanum melongena* L.) is a valuable vegetable cash crop mostly grown under plastic shed (Wang et al., 2015; Ghani et al., 2022a). Eggplant production in plastic shed heavily relies on mono-cropping systems. Like other crops, consecutive eggplant cultivation could occur mono-cropping obstacles, including an upsurge in autotoxins in the soil, which hampers plant growth and development, reduces resistance to harsh environmental conditions, and ultimately reduces plant yield and quality (Wang et al., 2015). One of the beneficial practices that ameliorate the negative impact of mono-cropping is crop rotation. To enhance productivity and optimize the profitability of a rotation system, rotated crops should be appropriately selected (Li et al., 2017; Ali et al., 2019). Diversified crop rotation during the fallow period mitigates the adverse effects of mono-cropping obstacles by sustaining the soil quality via nutrient deposit (St. Luce et al., 2020), greater SOM input (Ali et al., 2021), soil carbon sequestration (Song et al., 2018), and minimizing pest and disease attacks, particularly compared with mono-cropping systems (Ali et al., 2021; Ghani et al., 2022b). Different leafy vegetable plants have legacy effects on soil through root exudation, suppress soil-borne pathogens, and enhance soil fertility, improving plant growth and yield of the subsequent crop (Ali et al., 2019; Ghani et al., 2022b). Numerous studies have demonstrated yield advantages of crop rotation to subsequent crops, including tomato-onion, tomato-chrysanthemum, hairy vetch-eggplant, and cow pea-broccoli significantly improved tomato, eggplant, and broccoli yield compared with mono-cropping planting (Tian et al., 2009; Radicetti et al., 2016; Sánchez-Navarro et al., 2020). Furthermore, crop rotation also enhanced plant tolerances to different types of stresses (Gaudin et al., 2015).

Plants might be subjected to a wide range of external stresses during their growth period, including salt and heat stress and water deficit. However, crops grown in mono-cropping suffer from additional stresses such as evolving diseases and pests, various biotic and abiotic stresses, reduced soil physical-chemical characteristics, and the gradual buildup of root exudates in the soil. All of these stresses pose a constant threat to plant growth.

(Chen et al., 2011; Wang et al., 2015). Moreover, physiological and biochemical alterations following mono-cropping involve reduced growth underlined by a significant decline in photosynthetic rate. In addition, changes in soil pH adversely affect growth by declining photosynthesis and gas exchange parameters (Long et al., 2017). During stressful conditions, plants promote the production of reactive oxygen species (ROS) and cause significant damage to DNA, lipid peroxidation, the cell membrane, and proteins However, plants develop efficient adaptive strategies such as generating antioxidants, secondary metabolite, and osmolyte metabolism to overcome these stresses (Ahanger and Agarwal, 2017b; Ahanger et al., 2017; Ghani et al., 2022c). These biochemical pathways serve as an early signaling molecule of the plant's defense response to a variety of environmental stresses and as a secondary messenger for subsequent defense reactions (Yin et al., 2015; Chen et al., 2022). Increased activity of the antioxidant system and the metabolism of secondary metabolites avoid oxidative damage to membranes and other important macromolecules in the cell, hence keeping important mechanisms such as photosynthesis from being disrupted (Ahanger et al., 2017; Ahanger et al., 2018)

Previous studies have figure out the importance of crop rotation on pathogen suppression and soil stability (Thorup-Kristensen et al., 2012; Tian et al., 2013; Ali et al., 2019). More recently, we reported that leafy vegetables used as a crop rotation and their residue retention (above and below-ground biomass) significantly influenced soil nitrogen, microbial biomass, and soil enzymatic activity (Ghani et al., 2022b). However, limitation may still exist that dead roots and leaves were maintained in the field and well incorporated into the soil, and then influence on plant growth and development, as well as plant resistance-related enzymes. In this study, we postulated that executing different leafy vegetable plants would help to increase the eggplant's production capacity. Therefore, the aim of the study was to evaluate the capacities of different leafy vegetables, welsh onion, celery, non-heading Chinese cabbage, and leafy lettuce to alleviate mono-cropping obstacles of eggplant cultivation that generally occur due to mono-cropping, as well as to determine a sustainable vegetable cropping system to enhance eggplant production. Hence, the influence of different leafy vegetables on soil chemical properties, plant morphological and physiological observations, lipid peroxidation (MDA), and H_2O_2 level and, correlations between plant growth, physiology, and soil chemical properties were assessed.

2 Material and methods

2.1 Experimental site description and experimental design

Two years of field experiment under a plastic shed was conducted at the research station of Northwest A&F University Yangling, China. From March 2013 to November 2016, the eggplant was continuously cultivated under this plastic shed for four years. The eggplant was cultivated once a year with a mono-cropping

regime. The growing season of eggplant starts from 1st week of March to mid-November. The plastic shed's soil is sandy loam soil. The specific details of basic soil properties were reported previously by (Ghani et al., 2022b). After four years of consecutive eggplant cultivation, a rotational experiment was performed with four different winter leafy vegetables, including (I) welsh onion (Allium fistulosum L.), (II) celery (Apium graveolens L.), (III) non-heading Chinese cabbage (Brassica rapa L.), and (IV) leafy lettuce (Lactuca sativa L.) during the fallow period of eggplant from mid-November to 1st week of March. These leafy vegetables were harvested at the full leaf growth stage in 1st week of March, followed by the immediate planting of eggplant. The leafy vegetables root left over were mixed in the soil before eggplant seedlings were transplanted. With a factorial randomized complete block design, the eggplant seedlings were transplanted to the field in both years on the second week of March and harvested on 2nd week of November in both years. The experiment consisted of five different planting systems with three replications: fallow–eggplant (FE), welsh onion–eggplant (WOE), celery–eggplant (CE), no-heading Chinese cabbage–eggplant (NCCE), and leafy lettuce–eggplant (LLE) with three replications.

Each cropping system and fellow eggplant was assigned three plots and each plot consist of three beds. The size of each plot was 12.60 m^2 (3.6 m long . 3.5 m wide) and each bed was 4.20 m^2 (3.5 m long . 1.2 m wide). A thin, impermeable plastic sheet was inserted at a depth of 50 cm into the soil among experimental plots and extended to 5 cm above the ground, intended to prevent the passage of water and nutrients between plots and stop the interplay of various treatments. Three-week-old uniform eggplant seedlings (Solanummelongena L. Cv.Tai Kong Qie Wang) with 3 leaves were transplanted to the above-prepared beds; each bed consists of two rows. There were 7 seedlings in each row, and 14 in each bed, with 0.8 m distance between rows and 0.5 m between plants. Each year, before eggplant planting, each bed was fertilized with organic fertilizer (PengDiXin) at the rate of (50.65 kg ha^{-1}), "SaKeFu" (119.04 kg ha^{-1}). The detailed information is previously reported in (Ghani et al., 2022b). In addition, JinBa fertilizer was top-dressed (0.5kg/bed) according to local recommendations for vegetable production. No chemical fertilizers were used during the winter leafy vegetable growth period, and the same amount of irrigation, fertilization, and management practices were carried out throughout the experiment.

2.2 Measurement of morphological indices

To evaluate morphological traits at different growth phases, we randomly sampled three plants from each replication and 9 plants from each treatment. The growth phases included the first flowering, the first fruiting, the second flowering, and the second fruiting. A measuring tape was utilized to get an accurate reading of the plant height. The diameter of the stem was determined using an electronic vernier caliper. The eggplant's fresh weight (FW) was measured using an electronic scale after the eggplant was harvested, whereas the eggplant's dry weight (DW) was determined after oven drying at 70 °C for 72 hours.

2.3 Quantification of photosynthetic pigment and gas exchange parameters

Chlorophyll a, b, total chlorophyll, and carotenoids were measured by placing 0.5 g of fresh leaf tissue into a 25 mL glass tube with 20 mL 80% acetone for 48 hours in the dark after that, and absorbance was determined spectrophotometrically (UV-3802, UNICO, MDN, USA) at 645 nm, 663 nm, and 652 nm, respectively (Arnon, 1949).

Net photosynthetic rate (pN), stomatal conductance (gs), intercellular CO_2 concentration (Ci), and transpiration rate (E) were measured in the uppermost leaf by using a LI-6400 portable photosynthesis system (Li-Cor, Lincoln, NE, USA).

Measurement of the maximal photochemical efficiency (Fv/Fm), photosystem II (ΦPSII), non-photochemical quenching coefficient (NPQ) and, photochemical quenching (qP) were determined using modulated chlorophyll fluorometer (PAM-2000 chlorophyll fluorometer) after 20 min of dark adaptation. The recorded data were processed by PAM Win software.

2.4 Antioxidant enzymes assay

The eggplant leaves (0.5 g) were homogenized in a chilled 0.05 mM (pH 7.8) phosphate buffer containing 0.1% polyvinylpyrrolidone and 0.5 M ethylenediaminetetraacetic acid (EDTA). The homogenate was centrifuged at 12000 g for 15 min at 4°C, and the supernatant was used for enzyme analysis. To estimate superoxide dismutase (SOD) activity, we followed the method of (Dhindsa et al., 1981). The enzyme's ability to inhibit the photochemical reduction of nitroblue tetrazolium (NBT) was monitored at 560 nm.

The reaction mixture was prepared by adding 0.5 mL enzyme extract into 50 mL of 0.05 M phosphate buffer (pH 7.8), 28 mL guaiacol, and 19 mL 30% H_2O_2 (v/v) were mixed. The prepared reaction mixture of 3.5 mL was then transferred to a cuvette (1 cm) path length. An increase in absorbance was recorded at 470 nm wavelength over 3 min at 30-s intervals.

For the Catalase (CAT) activity, an enzyme extract of (0.1 mL) was added to 1.9 mL of reaction mixture containing phosphate buffer 200 mM (pH 7.0) and 1 mL of 0.3% H_2O_2. The enzyme activity was assessed by observing the decrease in absorbance at 240 nm for 2 min (Chance and Maehly, 1955).

Ascorbate peroxidase (APX) extraction was quantified by observing a reduction in absorbance due to the oxidation of ascorbic acid at 290 nm according to the method of (Nakano and Asada, 1981). The enzyme mixture consisted of 50 mM potassium phosphate buffer (pH 7.0), 0.2 mM EDTA, 0.5 mM ascorbic acid, 2 mM H_2O_2, and 100 μL enzyme extract.

Polyphenol oxidase (PPO) activity was assayed by measuring the initial increase in absorbance during the first 3 min of the reaction at 410 nm (Zheng et al., 2007). PAL activity was assessed according to the method of (Gao, 2006).

2.5 Measurement of oxidative stress biomarkers

The concentration of H_2O_2 was estimated following the method of (Velikova et al., 2000). For lipid peroxidation, the content of

malonaldehyde (MDA) was measured by incubating tissue extract with thiobarbituric acid (TBA) at boiling temperature (Dhindsa et al., 1981).

2.6 Assessment of protein content

The protein content was quantified using the method proposed by (Bradford, 1976) using BSA as a standard.

2.7 Measurement of soil chemical properties

Soil organic matter (SOM) was determined through the procedure described by (Walkley and Black, 1934). We utilized the alkali-hydrolyzed diffusion method to determine the amount of available nitrogen (AN) (Shi, 1996). The development of a complex that was blue in colour following the extraction of 0.5 M NaHCO3 at a pH of 8.5 was used to measure the amount of accessible phosphorus (AP) in the soil (Murphy and Riley, 1962). The following procedure was used to analyze the available K (Knudsen et al., 1983).

2.8 Statistical analysis

Data presented are the mean of three replicates and were statistically analyzed by two-way analysis of variance (ANOVA) as a 5×2 (treatment × year) factorial design for the experiment, and Tukey HSD test was used to analyze the mean separations among treatments at p < 0.05. Using origin software, a Pearson correlation was conducted to evaluate the relationship between plant growth, physiological traits, and soil chemical properties. All statistical analyses were performed with SPSS v.19.0 (SPSS Inc., Chicago, USA).

3 Results

3.1 Effect of different leafy vegetable cropping systems on morphological traits and chlorophyll pigments in eggplant leaves

Results illustrated in Table 1 showed that plant height and stem diameter were significantly ($P < 0.05$) affected by WOE, CE, NCCE and, LLE cropping systems as compared to FE. The highest plant height was observed for NCCE and LLE cropping systems than FE, and the observed increase was 18.34% and 16.15%, respectively. Regarding the growing stages, the eggplant height showed a rapid increase from 1st flowering to 2nd flowering stage and then slowed down at later stages. The maximum increase was recorded at 2nd fruiting stage (297.38%) than other growth stages, whereas the year factor exhibited a non-significant effect. Maximum stem diameter

was observed for CE as compared to FE, and the increase was 15.54%. Regarding growth stages, stem diameter increased gradually from 1st flowering to 2nd flowering. The maximum stem diameter was observed at 2nd fruiting stage as compared to other growth stages, and the increase was 45.32%. The year factor also had a significant effect, and the highest increase in stem diameter was observed in 2018 as compared to 2017 (Table 1).

Leafy vegetable cropping systems significantly stimulated the chlorophyll pigments. Maximum chlorophyll a content was observed for NCCE as compared to FE rotation, and the increase was 13.48% (Table 1). With respect to growth stages, maximum content of chlorophyll a was recorded at 1st flowering stage as compared to other growth stages, and the increase was 41.27%, and the minimum chlorophyll a was observed at 2nd fruiting stage. Year factor was also significant, and maximum chlorophyll a was observed in 2018. The interaction effect of three factors (Y × T × S) on chlorophyll a was also significant. Winter leafy vegetables significantly affected chlorophyll b pigment. The highest increase (35.27%) in chlorophyll b content was observed for NCCE as compared to FE rotation. Regarding growing stages, maximum chlorophyll b was observed at 1st flowering stage, and an increase was 52.24%, and minimum chlorophyll b was observed at 2nd fruiting stage. Year factor was also significant, and maximum chlorophyll b was observed in 2018. The interaction effect of three factors (Y × T × S) on chlorophyll b was also significant (Table 1).

It is shown in Table 1 that chlorophyll a and b were significantly ($P < 0.05$) affected by leafy vegetable rotation systems. It was shown that the highest chlorophyll ab were recorded for NCCE, followed by CE as compared to FE, and an increase was 9.20% and 7.69%, respectively. Chlorophyll ab were observed maximum at 1st flowering stage, and an increase was 33.27%, and the minimum increase (24.96%) was observed at the 2nd fruiting stage. Year factor was also significant, and maximum chlorophyll ab was observed in 2018. The interaction effect of three factors (Y × T × S) on chlorophyll ab was also significant.

The effect of different leafy vegetable species on carotenoid content is illustrated in Table 1. Leafy vegetables significantly ($P < 0.05$) enhanced the carotenoid contents compared to FE, and a significant effect of NCCE was observed on carotenoid content. Maximum carotenoid content was observed for NCCE as compared to FE, and the increase was 26.7%. The growth stage also had a significant effect on carotenoid content. The highest increase in carotenoid content was observed at the 1st flowering stage, and the lowest increase was observed at the 2nd fruiting stage. The year factor was significant, and the highest carotenoid content was observed in 2018.

3.2 Effect of different leafy vegetable cropping systems on gas exchange parameters

Leafy vegetable cropping systems significantly ($P < 0.05$) affected gas exchange parameters (Table 2). pN was higher for CE

TABLE 1 Effect of different leafy vegetable rotation systems on morphological indexes and photosynthetic pigments of eggplant at different growth stages during the years 2017 and 2018.

Treatment	Plant height (cm)	Stem diameter (mm)	Chlorophyll a (mg g^{-1} FW)	Chlorophyll b (mg g^{-1} FW)	Chlorophyll ab (mg g^{-1} FW)	Carotenoids (mg g^{-1} FW)
Year						
2017	111.1 ± 48.57a	14.200 ± 2.64b	17.18 ± 2.44b	6.77 ± 2.23b	24.42 ± 2.90b	2.96± 1.35b
2018	112.2 ± 48.32a	14.771 ± 2.79a	18.16 ± 2.46a	7.63 ± 2.02a	25.63 ± 2.66a	3.69 ± 2.66a
Leafy vegetables						
FE	100.3 ± 43.95c	13.941 ± 2.33c	16.84 ± 2.36d	6.18 ± 1.95e	23.90 ± 3.05d	2.87± 1.23d
WOE	107.1 ± 41.35b	15.196 ± 2.41b	16.98 ± 2.34d	6.68 ± 1.96d	24.49 ± 2.82c	3.19 ± 1.46c
CE	115.8 ± 51.16a	16.107 ± 2.30a	18.00 ± 2.27b	7.73 ± 2.01b	25.74 ± 2.68a	3.57 ± 1.51ab
NCCE	118.7 ± 52.01a	14.906 ± 2.62b	19.11 ± 2.74a	8.36 ± 1.99a	26.10 ± 2.30a	3.62 ± 1.49a
LLE	116.5 ± 50.31a	12.275 ± 2.32d	17.41 ± 2.04c	7.05 ± 2.22c	24.89 ± 2.77b	3.40± 2.77bc
Stage						
1st flowering	39.81 ± 3.69d	12.145 ± 1.30c	20.81 ± 1.29a	9.80 ± 1.03a	28.48 ± 0.88a	5.50 ± 0.62a
1st fruiting	97.95 ± 6.37c	12.509 ± 1.34c	18.57 ± 0.88b	8.10 ± 0.95b	26.14 ± 0.85b	3.40 ± 0.82b
2nd flowering	150.8 ± 11.34b	15.636 ± 1.68b	16.56 ± 1.02c	6.22 ± 2.01c	24.10 ± 1.19c	2.47 ± 0.28c
2nd fruiting	158.2 ± 12.34a	17.650 ± 1.66a	14.73 ± 0.98d	4.68 ± 1.99d	21.37 ± 1.50d	1.94 ± 0.44d
F-test						
Year (Y)	ns	***	***	***	***	***
Leafy vegetable (LV)	***	***	***	***	***	***
Stage (S)	***	***	***	***	***	***
Y × LV × S	ns	ns	*	**	*	*

Data are presented as means with standard deviation (n= 9). Different letters show significant difference at p<0.05 level. FE, fellow eggplant; WOE, welsh onion-eggplant: CE, celery-eggplant; NCCE, non-heading Chinese cabbage-egglant; LLE, leafy lettuce-eggplant. *p < 0.05; **p < 0.01; ***p < 0.001; ns, non-significant.

and NCCE than FE, with an increase of 15.28% and 11.18%, respectively. Similarly, growth stages also had a significant effect on net photosynthesis. pN increased at all stages, and the maximum increase was observed at 1st flowering stage, which was 39.94%, while the minimum increase was observed at 2nd fruiting stage, which was 28.54%. Year factor was also significant, and maximum pN was observed in 2018. The interaction effect of three factors (Y × T × S) on pN was also significant (Table 2). In addition, leafy vegetable species had a significant impact on Ci. Ci was increased in all vegetable species as compared to FE. NCCE and CE exhibited higher Ci than FE, and the increase was 11.15% and 10.17%, respectively. The Ci increased at all growth stages, with the highest values recorded at 1st flowering stage. The year factor had a non-significant effect on Ci (Table 2).

Leafy vegetable rotations also had a significant (P < 0.05) effect on gs and E. The highest increase in gs and E was observed in NCCE as compared to FE, and the increase was 50% and 42.47%, respectively (Table 3). Similarly, growth stages also significantly affected gs and E. Maximum gs was observed at the 1st flowering stage with an increase of 375%, and maximum E was observed at the 1st flowering stage 11.11%. The year factor had a non-significant

effect on gs. Whereas, for E year factor was significant and maximum E was observed in 2018 (Table 2).

3.3 Effect of different leafy vegetable cropping systems on chlorophyll fluorescence parameters

Crop rotation with leafy vegetable cropping systems significantly (P < 0.05) affected Fv/Fm, ΦPSII, NPQ, and qp (Table 3). Among different leafy vegetable rotations, the maximum increase in Fv/Fm, ΦPSII, and qP was observed for NCCE cropping system as compared to FE, and the increase was 8.57, 9.23, 25.92, and 14.47%, respectively. At different growth stages, maximum Fv/Fm, ΦPSII, and qP were observed at 1st flowering stage, and the increase was 8.69%, 9.23%, and 9.09%, respectively. However, NPQ was higher in FE than other leafy vegetables and increased by 6.89% compared to other leafy vegetables (Table 4). The year factor was also significant for chlorophyll fluorescence parameters, and maximum Fv/Fm, ΦPSII, and NPQ were observed in 2017, and qP was observed maximum in 2018 (Table 3).

TABLE 2 Effect of different leafy vegetable rotation systems on gas exchange parameters of eggplant at different growth stages during the year 2017 and 2018.

Treatment	(pN) ($\mu mol\ m^{-2}\ s^{-1}$)	Ci ($\mu mol\ mol^{-1}$)	E ($mmol\ m^{-2}\ s^{-1}$)	gs ($mol\ m^{-2}\ s^{-1}$)
Year				
2017	26.50 ± 3.40a	281.0 ± 30.04a	0.029 ± 0.171a	1.38 ± 0.21a
2018	25.37 ± 3.55b	280.2 ± 19.39a	0.008 ± 0.006b	1.34 ± 0.22a
Leafy vegetables				
FE	23.88 ± 2.90d	263.5 ± 26.37d	0.006 ± 0.006c	1.13 ± 0.15d
WOE	25.52 ± 3.09c	271.7 ± 25.06c	0.007 ± 0.006b	1.28 ± 0.13c
CE	27.53 ± 3.59a	290.3 ± 22.29ab	0.008 ± 0.007a	1.43 ± 0.18b
NCCE	26.55 ± 3.76b	292.9 ± 19.22a	0.009 ± 0.007a	1.61 ± 0.13a
LLE	26.18 ± 3.11b	284.5 ± 19.42b	0.007 ± 0.006b	1.36 ± 0.13c
Stage				
1st flowering	30.13 ± 1.98a	308.0 ± 8.98a	0.019 ± 0.002a	1.40 ± 0.21a
1st fruiting	27.32 ± 1.58b	287.5 ± 15.60b	0.003 ± 0.001c	1.40 ± 0.15a
2nd flowering	24.75 ± 1.24c	269.1 ± 15.45c	0.004 ± 0.001b	1.38 ± 0.22a
2nd fruiting	21.53 ± 1.23d	257.8 ± 23.32d	0.004 ± 0.001b	1.26 ± 0.24b
F-test				
Year (Y)	***	ns	***	ns
Leafy vegetable (LV)	***	***	***	***
Stage (S)	***	***	***	***
Y × LV × S	ns	ns	**	ns

Main effect due to treatment (crop rotation), sampling year and their interaction was analyzed by Two-way ANOVA. Data are presented as means with standard deviation (n= 9). Different letters show significant difference at p<0.05 level. pN, net photosynthesis; Ci, internal CO_2 rate; E, transpiration rate. gs, stomatal conductance; FE, fellow eggplant; WOE, welsh onion-eggplant: CE, celery-eggplant; NCCE, non-heading Chinese cabbage-egglant. ** $p < 0.01$; *** $p < 0.001$; ns, non-significant.

3.4 Effect of different leafy vegetable cropping systems on antioxidant system and soluble protein

Results for the antioxidant enzymes observed are depicted in Table 4. The results revealed that WOE, CE, NCCE and, LLE cropping systems used as a crop rotation enhanced the antioxidant enzyme of eggplant. Among different winter leafy vegetable species, NCCE rotation had shown maximum activity of SOD, POD, PAL, PPO, APX, and CAT as compared to FE, and the increase was 38.36%, 53.68%, 27.78%, 21.44%, 18.90%, and 28.00%, respectively. Furthermore, with respect to different growth stages, various enzymatic activities were observed maximum at 2nd fruiting stage than other growth stages, and a significant increase of SOD (47.01%), POD (36.55%), PAL (56.43%), PPO (77.48%), APX (42.22%) and CAT (27.70%) was observed (Table 4). Year factor was also significant, and maximum SOD, POD, PAL, APX, and CAT were observed in 2017, whereas maximum PPO was observed in 2018. The interaction effect of three factors (Y × T × S) on SOD, PAL, PPO, APX, and CAT was also significant, whereas POD activity was non-significant.

Leafy vegetable cropping systems significantly influenced the soluble protein content of eggplant leaves (Table 4). A maximum increase in

soluble protein content was observed for NCCE (26.37%) and CE (25.41%) as compared to FE. A higher amount of soluble protein was observed at different growing stages at 2nd fruiting stage than in other growth stages. The year factor was also significant, and higher soluble protein content was observed in 2018 than in 2017 (Table 4).

3.5 Effect of different leafy vegetable cropping systems on MDA and H_2O_2

Using different Leafy vegetable cropping systems as a crop rotation significantly ($P < 0.05$) reduced MDA and H_2O_2 concentration compared to FE (Table 5). A higher reduction in MDA concentration was recorded for CE and NCCE as compared to FE, and an increase was 33.09% and 32.84%, respectively. The year factor was also significant, and a higher concentration was observed in 2018 than in 2017. Similarly, H_2O_2 was also lower in all leafy vegetable treatments as compared to FE. Compared to FE, the maximum reduction was observed under NCCE (28.52%) and CE (26.59%). Regarding different growth stages, a higher reduction in MDA and H_2O_2 concentration was recorded at the 2nd fruiting stage, and a higher reduction of 60.45% and 51.16% in MDA and

TABLE 3 Effect of different leafy vegetable rotation systems on chlorophyll fluorescence parameters of eggplant at different growth stages during the year 2017 and 2018.

Treatment	Fv/Fm	ΦPSII	NPQ	qP
Year				
2017	0.73 ± 0.03a	0.70 ± 0.03a	0.34 ± 0.02a	0.79 ± 0.03b
2018	0.71 ± 0.03b	0.66 ± 0.03b	0.26 ± 0.02b	0.83 ± 0.06a
Leafy vegetables				
FE	0.70 ± 0.02d	0.65 ± 0.03d	0.34 ± 0.04a	0.76 ± 0.02e
WOE	0.71 ± 0.02c	0.68 ± 0.02b	0.29 ± 0.03c	0.78 ± 0.02c
CE	0.74 ± 0.02b	0.70 ± 0.02a	0.32 ± 0.04b	0.85 ± 0.05b
NCCE	0.76 ± 0.02a	0.71 ± 0.02a	0.27 ± 0.04d	0.87 ± 0.05a
LLE	0.70 ± 0.02d	0.67 ± 0.02c	0.29 ± 0.03c	0.78 ± 0.02d
Stage				
1st flowering	0.75 ± 0.02a	0.71 ± 0.03a	0.31 ± 0.05a	0.84 ± 0.05a
1st fruiting	0.73 ± 0.02b	0.69 ± 0.03b	0.31 ± 0.04b	0.82 ± 0.05b
2nd flowering	0.71 ± 0.02c	0.67 ± 0.02c	0.30 ± 0.04c	0.80 ± 0.05c
2nd fruiting	0.69 ± 0.02d	0.65 ± 0.02d	0.29 ± 0.04d	0.77 ± 0.05d
F-test				
Year (Y)	***	***	***	***
Leafy vegetable (LV)	***	***	***	***
Stage (S)	***	***	***	***
Y × LV × S	ns	ns	ns	ns

Data are presented as means with standard deviation (n= 9). Different letters show significant difference at p<0.05 level. Fv/Fm, photochemical efficiency; ΦPSII, photosystem II; NPQ, non-photochemical quenching coefficient; qP, photochemical quenching. FE, fellow eggplant; WOE, welsh onion-eggplant: CE, celery-eggplant; NCCE, non-heading Chinese cabbage-eggplant. *** p<0.001; ns, non-significant.

H_2O_2 respectively was recorded at the 2nd fruiting stage compared to the 1st flowering stage (Table 5).

3.6 Effect of different leafy vegetable cropping systems on soil chemical properties

Different leafy vegetable cropping systems used as a crop rotation showed a significant ($P < 0.05$) impact on soil chemical properties such as pH, EC, SOM, and soil available nutrients in both years (Table 6). A maximum increment in soil pH was recorded in NCCE by (8.01%) and CE (6.61%) compared with FE. The highest increase in pH was observed at 2nd flowering stage as compared to other stages (Table 6). The year factor was also significant, and a higher concentration was observed in 2018 than in 2017. The interaction effect of three factors (Y × T × S) on pH was also significant (Table 6). Soil EC showed a downward trend after the inclusion of leafy vegetable cropping systems (Table 6). WOE and NCCE exhibited the maximum reduction in EC. The highest increase in EC was observed at 1st flowering stage as compared to other stages (Table 6). The year factor was also significant, and a higher concentration was observed in 2018 than in 2017. Similarly,

leafy vegetables significantly impacted AN, AP, and SOM. The NCCE exhibited a maximum increment in AN (24.31%), AP (14.81%), and SOM (26.37%) compared with the FE. The year factor was significant, with the highest increment in AP and SOM in 2018 than in 2017 (Table 6). The highest increase was observed at 2nd fruiting stage as compared to other stages (Table 6). However, AK was higher in CE, which increased AK by 9.78% compared to FE. The year factor was significant, with the highest increment in AP in 2017 than in 2018 (Table 6). The highest increase was observed at 2nd fruiting stage as compared to other stages.

3.7 Effect of different leafy vegetable cropping systems on eggplant fresh and dry biomass

Leafy vegetable cropping systems significantly ($P < 0.05$) increased eggplant's fresh and dry weight over FE. Both years, the highest increment in eggplant biomass was observed in NCCE and, CE rotation (Table 7). However, the highest enhancement was observed in 2018, and the observed increase was 73.05% in NCCE and 62.06% in CE rotation. The year factor also had a significant effect; the highest increase in fresh weight was observed in 2018

TABLE 4 Effect of different leafy vegetable rotation systems on the antioxidant system and soluble protein of eggplant at different growth stages during the year 2017 and 2018.

Treatment	SOD activity (U g^{-1} FW h^{-1})	POD activity (U mg^{-1} protein min^{-1})	PAL (A290 g^{-1} h^{-1})	PPO (0.001ΔA min^{-1})	APX activity Mm/(g. min)	CAT activity (U mg^{-1} protein min^{-1})	Soluble protein (mg g^{-1})
Year							
2017	615.8 ± 175.9b	18.76 ± 4.51b	2758 ± 819.9b	368.5 ± 222.5a	7.96 ± 1.52b	25.89 ± 5.24b	15.82 ± 2.88b
2018	659.3 ± 157.5a	19.29 ± 4.71a	2820 ± 799.6a	324.0 ± 187.0b	8.66 ± 2.02a	27.68 ± 4.16a	16.30 ± 2.81a
Leafy vegetables							
FE	537.0 ± 151.3e	15.20 ± 2.70d	2462 ± 723.6e	307.8 ± 175.9d	7.51 ± 1.80e	23.57 ± 4.05d	14.14 ± 2.14c
WOE	560.5 ± 154.9d	16.39 ± 2.91c	2642 ± 761.4d	332.1 ± 192.0c	7.87 ± 1.90d	25.11 ± 4.03cd	15.26 ± 1.77b
CE	692.3 ± 142.6b	20.38 ± 3.77b	2957 ± 768.2b	369.3 ± 220.9a	8.73 ± 1.62b	28.17 ± 4.06ab	17.68 ± 3.33a
NCCE	743.0 ± 153.0a	23.36 ± 4.51a	3146 ± 819.3a	373.8 ± 227.6a	8.93 ± 1.72a	30.17 ± 4.72a	17.87 ± 2.51a
LLE	654.7 ± 143.7c	19.80 ± 3.70b	2737 ± 795.3c	348.0 ± 205.7b	8.50 ± 1.67c	26.90 ± 4.23bc	15.36 ± 2.19b
Stage							
1st flowering	438.0 ± 57.51d	15.03 ± 2.14d	1635 ± 191.9d	130.4 ± 6.83d	6.28 ± 0.80d	21.81 ± 3.81c	13.00 ± 1.56d
1st fruiting	594.0 ± 125.5c	17.09 ± 3.16c	2686 ± 273.1c	161.3 ± 18.44c	7.63 ± 0.63c	25.65 ± 3.36b	15.52 ± 1.92c
2nd flowering	690.5 ± 107.6b	20.30 ± 3.37b	3082 ± 318.4b	514.0 ± 50.55b	8.46 ± 0.48b	29.51 ± 3.19a	17.43 ± 2.03b
2nd fruiting	827.6 ± 45.81a	23.69 ± 4.06a	3753 ± 240.3a	579.2 ± 69.96a	10.87 ± 0.96a	30.17 ± 3.44a	18.30 ± 2.39a
F-test							
Year (Y)	***	*	***	***	***	***	*
Leafy vegetable (LV)	***	***	***	***	***	***	***
Stage (S)	***	***	***	***	***	***	***
Y × LV × S	**	ns	***	***	***	ns	ns

Data are presented as means with standard deviation (n= 9). Different letters show significant difference at p<0.05 level. SOD, super oxidase; POD, peroxidase; PAL, phenylalanine ammonia–lyase; PPO, polyphenol oxidase; APX, ascorbate peroxidase: CAT, catalase. FE, fellow eggplant; WOE, welsh onion-eggplant: CE, celery-eggplant; NCCE, non-heading Chinese cabbage-egglant. *p < 0.05; **p < 0.01; ***p < 0.001; ns, non-significant.

TABLE 5 Effect of different leafy vegetable rotation systems on the oxidative markers of eggplant at different growth stages during the year 2017 and 2018.

Treatment	MDA content (nmol g^{-1} FW)	H$_2$O$_2$ (µmol g^{-1} FW)
Year		
2017	28.00 ± 12.09 b	21.68 ± 6.47a
2018	29.59 ± 11.89a	21.36 ± 6.12a
Leafy vegetables		
FE	36.35 ± 12.53a	25.87 ± 7.23a
WOE	29.30 ± 10.91b	22.81 ± 6.39b
CE	24.32 ± 10.29c	18.99 ± 4.33d
NCCE	24.41 ± 10.59c	18.49 ± 4.35d
LLE	29.61 ± 11.45b	21.43 ± 5.60c
Stage		
1st flowering	17.86 ± 4.01d	13.37 ± 1.60d
1st fruiting	21.58 ± 5.39c	20.73 ± 2.96c
2nd flowering	30.46 ± 5.90b	24.60 ± 4.19b
2nd fruiting	45.30 ± 7.17a	27.38 ± 4.34a
F-test		
Year (Y)	*	ns
Leafy vegetable (LV)	***	***
Stage (S)	***	***
Y × LV × S	ns	ns

Data are presented as means with standard deviation (n= 9). Different letters show significant difference at p<0.05 level. MDA: malondialdehyde; H$_2$O$_2$: hydrogen peroxide. FE, fellow eggplant; WOE, welsh onion-eggplant: CE, celery-eggplant; NCCE, non-heading Chinese cabbage-egglant. * p<0.05; *** p<0.001; ns, non-significant.

compared to 2017 (Table 7). Similarly, dry weight was highest in NCCE and CE in both years, which increased by 51.01% in CE and 71.45% in NCCE over FE in 2018.

3.8 Correlation between plant physiological, biochemical indexes, and soil chemical properties

The correlation between various physiological, biochemical, and soil chemical characteristics was examined using the Pearson correlation (Figure 1). Based on Pearson correlation PFB, PDB showed a highly significant positive correlation with SOM, AN, AP, and AK. However, these traits showed a highly negative correlation with NPQ, MDA, HD, and EC. Similarly, plant enzymes such as SOD, POD, CAT, PAL, PPO, and APX showed a highly negative correlation with NPQ, MDA, HD, and EC. SOD, POD, CAT, PAL, PPO, and APX positively correlated with SOM. CAT exhibited a positive correlation with AK, while PPO showed a positive correlation with AK.

4 Discussion

Our results demonstrated that two years of vegetable cover crops used as a crop rotation significantly impacted soil pH (Table 6). It is likely due to the incorporation of dead roots in the soil after harvesting leafy vegetables. Aschi et al. (2017) reported that the residues' chemical properties could change the soil pH, primarily due to their alkalinity and nitrogen content. Soil organic carbon (SOC) is considered a good soil quality index Ghani et al. (2022d) because it affects soil fertility. Many studies have reported that in crop rotation, below-ground rhizodeposits and root addition are the key factors of carbon accumulation in the soil, accounting for up to 75% of soil organic matter (Jones et al., 2009; Ghani et al., 2022d). Different types of crop stalks, twigs, dead roots as well as fallen leaves are important sources of nutrients Aschi et al. (2017), while leafy vegetables used as crop rotation can scavenge nutrients from the soil, store it in their residues and return it to the soil for the next crop through root decomposition and improved SOM (Dorissant et al., 2022; Ghani et al., 2022b). Kong and Six (2012) reported that living cover crops significantly improved soil organic matter compared to bare fallow due to the rhizodeposition of low molecular carbon into the soil. The rhizodeposition rate decreases with plant age, but the addition of mature roots into the soil as residue acts as a microbial substrate, thus increasing MBC as well as soil organic matter (Chahal and Van Eerd, 2020). In line with this concept, NCCE and CE modify the soil environment (Table 6) by improving soil nutrients and SOM through root exudates and dead roots, which were available for eggplant. Our results align with the findings of Ali et al. (2021), where different winter leafy vegetables cover crops used as a crop rotation can enhance soil nutrient availability and SOM by incorporating plant and root residue.

It was observed that eggplant grown on soils after the leafy vegetable crop rotation exhibited increased growth in terms of plant height, stem diameter, and greater biomass (Tables 1, 7) and reflected in significantly increased yield (Ghani et al., 2022b) compared with mono-cropping for two years. Higher eggplant growth and biomass production directly correlated to improved soil chemical properties, evident from the positive correlation between plant growth and soil chemical parameters (Figure 1). Similar findings were reported by D'Acunto et al. (2018), where pea-maize crop rotation improved maize biomass which was interlinked with different soil properties.

Furthermore, increased growth and yield in eggplant-leafy vegetable cropping systems were correlated with enhanced photosynthesis and gas exchange, which was confirmed by the positive correlation between eggplant growth and photosynthetic parameters (Figure 1). Earlier, it has been reported that mono-cropping or intercropping systems negatively influence the photosynthesis and growth of plants (Yao et al., 2017). Mono-cropping results in the depletion of soil microbial population and mineral status. Mono-cropping of cucumber reduces growth and yield by declining beneficial microbial populations in the soil, and crop rotation significantly increases the yield (Wu and Wang, 2007). Reduced photosynthesis directly results from the restricted

TABLE 6 Effect of different leafy vegetable rotation systems on soil chemical properties of eggplant at different growth stages during the year 2017 and 2018.

Treatment	pH	EC ($\mu s \cdot cm^{-1}$)	AN ($mg\ kg^{-1}$)	AP ($mg\ kg^{-1}$)	AK ($mg\ kg^{-1}$)	SOM ($g\ kg^{-1}$)
Year						
2017	7.49±0.17b	162.14±60.43a	125.97 ± 3.39a	34.37 ± 0.57b	360.58 ± 9.43a	21.08 ± 0.53b
2018	7.51±0.22a	160.36±71.71b	123.19 ± 3.70b	37.67 ± 0.65a	351.32 ± 9.97b	22.93 ± 0.35a
Leafy vegetables						
FE	7.11±0.09d	288.58±11.42a	113.79 ± 2.99c	34.42 ± 1.20b	338.53 ± 3.63d	19.30 ± 0.36e
WOE	7.56±0.06bc	113.80±5.81e	116.13 ± 3.09c	35.19 ± 1.16b	355.06 ± 2.32c	22.06 ±0.0.49c
CE	7.58±0.05b	126.33±7.47c	136.06 ± 3.59b	35.70 ± 1.25b	371.64 ± 2.90a	23.08 ± 0.28b
NCCE	7.68±0.04a	120.59±5.98d	141.46 ± 3.76a	39.52 ± 1.42a	363.89 ±2.23b	24.39 ± 0.33a
LLE	7.57±0.06bc	154.42±4.98b	115.45 ± 3.11c	35.29 ± 1.50b	350.61 ± 1.58c	21.20 ± 0.17d
Stage						
1st flowering	7.49±0.18b	162.32±20.32a	123.77 ± 1.98b	33.75 ± 2.92bc	343.94 ± 2.57d	21.65 ± 0.56c
1st fruiting	7.50±0.20ab	161.58±21.76a	129.79 ± 1.58a	35.49 ± 2.62c	354.35 ± 2.70c	21.54 ± 0.20c
2nd flowering	7.51±0.23a	159.59±22.40b	119.69 ± 1.24c	36.46 ± 2.92b	359.71 ± 2.80b	22.16 ± 0.25b
2nd fruiting	7.50±0.24ab	159.48±22.57b	125.05 ± 1.23b	38.40 ± 2.69a	365.80± 2.61b	22.67 ± 0.44ba
F-test						
Year (Y)	***	***	*	*	*	***
Leafy vegetable (LV)	***	***	***	***	***	***
Stage (S)	*	***	***	***	***	***
Y × LV × S	***	ns	ns	ns	ns	ns

Data are presented as means with standard deviation (n= 9). Different letters show significant difference at p<0.05 level. AN, available nitrogen; AP, available phosphorus; AK, available potassium; SOM, soil organic matter; FE, fellow eggplant; WOE, welsh onion-eggplant: CE, celery-eggplant; NCCE, non-heading Chinese cabbage-egglant. * p<0.05; *** p<0.001; ns, non-significant.

stomatal conductance and internal CO_2 concentrations in mono-cropped eggplant. It is believed that increased stomatal conductance leads to the maintenance of CO_2, reflected in increased photosynthetic rate (Ahanger et al., 2018). Reports discussing the role of leafy vegetables-eggplant cropping systems in the protection of the photosynthetic system, growth, and yield are not available. The present investigation shows the protection of available mineral status and soil health due to crop rotation with leafy vegetables. Reduced photosynthesis in the mono-cropping or intercropping system results in declined production of energy, carbohydrate

TABLE 7 Effect of different leafy vegetable rotation systems on fresh and dry weight of eggplant during the year 2017 and 2018.

Treatment	Fresh Weight (g)			Dry Weight (g)		
	2017	2018	Means	2017	2018	Means
FE	351.41±8.18e	342.13±11.48e	346.77±6.89e	140.70±2.95fg	127.66±4.47g	134.18±10.58e
WOE	448.74±9.00d	474.12±8.05cd	461.43±10.94c	163.53±12.03de	174.12±5.69cd	168.83±4.14c
CE	521.35±12.39bc	454.47±15.86ab	537.91±5.49b	187.88±4.06c	192.79±4.27bc	190.34±7.12b
NCCE	553.35±10.49ab	592.08±11.22a	572.72±11.46a	208.40±5.14ab	218.88±5.34a	213.64±7.90a
LLE	370.74±9.03ef	389.28±9.49e	380.01±7.05d	148.23±4.00ef	153.28±3.89ef	150.76±5.30d
Year means	449.12±21.61b	470.42±25.65a		169.75±6.89a	173.35±8.58a	
Tukey HSD test	Treatment	Year	Interaction	Treatment	Year	Interaction
	***	***	NS	***	NS	*

Data are presented as means with standard deviation (n= 3). Different letters show significant difference at p<0.05 level. FE, fellow eggplant; WOE, welsh onion-eggplant: CE, celery-eggplant; NCCE, non-heading Chinese cabbage-egglant. * p<0.05; *** p<0.001; ns, non-significant.
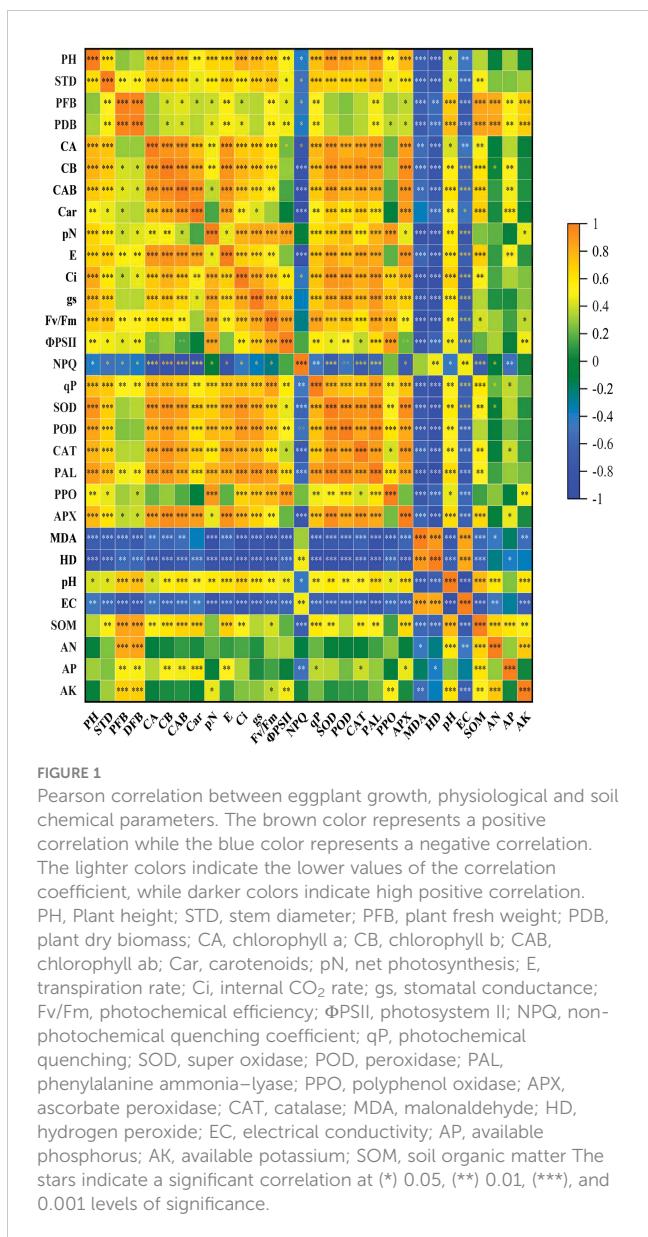
FIGURE 1

Pearson correlation between eggplant growth, physiological and soil chemical parameters. The brown color represents a positive correlation while the blue color represents a negative correlation. The lighter colors indicate the lower values of the correlation coefficient, while darker colors indicate high positive correlation. PH, Plant height; STD, stem diameter; PFB, plant fresh weight; PDB, plant dry biomass; CA, chlorophyll a; CB, chlorophyll b; CAB, chlorophyll ab; Car, carotenoids; pN, net photosynthesis; E, transpiration rate; Ci, internal CO_2 rate; gs, stomatal conductance; Fv/Fm, photochemical efficiency; ΦPSII, photosystem II; NPQ, non-photochemical quenching coefficient; qP, photochemical quenching; SOD, super oxidase; POD, peroxidase; PAL, phenylalanine ammonia−lyase; PPO, polyphenol oxidase; APX, ascorbate peroxidase; CAT, catalase; MDA, malonaldehyde; HD, hydrogen peroxide; EC, electrical conductivity; AP, available phosphorus; AK, available potassium; SOM, soil organic matter The stars indicate a significant correlation at (*) 0.05, (**) 0.01, (***), and 0.001 levels of significance.

metabolism and chlorophyll production (Table 2), (Su et al., 2014). Crop rotation improves soil physical and chemical environment (Table 6), including water holding capacity and aeration, and ultimately increases plant growth attributes like root growth and nutrient foraging. In the wheat-peanut crop rotation system, increased nitrogen uptake and allocation resulted in greater chlorophyll synthesis and photosynthesis rate (Liu et al., 2019). Moreover, increased pigment synthesis and photosynthesis in the crop rotation system were linked with improved PSII functioning. Relative to FE, Fv/Fm, ΦPSII, and qP increased in all cropping systems while NPQ exhibited a reduction (Table 3). Photosystem functioning was considerably enhanced during both experimental seasons due to crop rotation with WOE, CE, NCCE, and LE.net Increased fluorescence parameters reflect photosynthetic regulation through non-stomatal modulations (Ahanger et al., 2020). The present study envisaged that both stomatal and non-stomatal enhancements in the different cropping systems influenced the

growth and yield performance of eggplant. Increased access to external CO_2 and reduced accumulation of toxic radicals in intercropping systems significantly contributed to the functioning of photosynthetic machinery (Yao et al., 2017).

Increased growth and photosynthesis in plants raised in the crop rotation system are due to reduced oxidative damage in them (Table 5), resulting in a significant enhancement in the structural and functional integrity of membranes and cells (Liu et al., 2019). Reduced generation of reactive oxygen species prevents the oxidative effects on the membranes reflecting in the maintenance of plant functioning (Ahanger and Agarwal, 2017b). In the present study, eggplant grown on soils following crop rotation with different leafy vegetables also exhibited reduced lipid peroxidation due to a significant decline in the accumulation of hydrogen peroxide during both years. Our findings were further supported by a negative correlation among growth, photosynthesis, and oxidative markers (Figure 1). The declining oxidative effects of mono-cropping due to the growth of leafy vegetables were observed due to up-regulation of APX and CAT activities in them during both years and at both developmental stages (Table 4). Relative to FE, activities of APX and CAT increased in seedlings grown after leafy vegetable crops, with the maximal increase in plants grown in NCCE system at both developmental stages. APX and CAT act on the same substrate, i.e., hydrogen peroxide but at different sites, with APX eliminating excess hydrogen peroxide from chloroplast while CAT from the cytosol (Ahmad, 2010). Increased APX activity due to crop rotation with leafy vegetables strengthened the key radical scavenging pathways, including ascorbate-glutathione in chloroplasts leading to increased protection of major cellular pathways, including photosynthesis (Ahanger and Agarwal, 2017a; Ghani et al., 2022a). Up-regulation of the activities of APX due to crop rotation prevents the formation of toxic hydroxyl radicals by assisting in the maintenance of redox homeostasis and the electron donors, including ascorbate and glutathione (Khan and Khan, 2014). Greater ascorbate and glutathione content due to improved ascorbate-glutathione functioning maintains the electron transport in chloroplasts and mitochondria (Nahar et al., 2016).

In addition to this protein content of eggplant increased significantly due to crop rotation with leafy vegetables. During both years, maximal protein content was reported at the second fruiting and second flowering stages. During both years of experimentation, the influence of leafy vegetable crop rotation on eggplant protein content slowly increased from 2017 to 2018 and showed an increasing trend with the developmental stage (Table 6), indicating the development of beneficial proteins under a crop rotation system. Proteins form an important nutritional component of plants, particularly in vegetables, and help plants maintain growth and development under different growth conditions (Lee and Yaffe, 2016). Proteins assist in signaling and maintaining development from seed germination to flowering. Plants have specific proteins maintaining key cellular functioning like photosynthesis, signaling, and response elicitation. Intercropping and crop rotation systems have been proposed to influence plant development by modulating physiology and biochemistry and reducing disease incidence; however, the effect has been reported to be species-dependent (Crow et al., 2000; Ghani et al., 2019a).

Such beneficial effects of crop rotation with leafy vegetables can be due to the significant decline in the accumulation of allelopathic compounds within the eggplant rhizosphere, thereby declining the growth through auto-allelopathy (Cheng and Cheng, 2015). However, eggplant crop rotation with leafy vegetables may have reduced the accumulation of allelochemicals and improved the synthesis of some specific proteins. Proteins mediate specific defense pathways in plants (Chuang et al., 2016). In addition, the activity of PAL is stimulated significantly due to crop rotation with vegetable crops.

PAL regulates the synthesis of secondary metabolites in plants. Increased PAL activity has been reported to contribute to greater stress tolerance by enhancing antioxidant potential (Ahanger and Agarwal, 2017b). Increased PAL activity due to crop rotation with leafy vegetables may improve eggplant metabolite content with significant health benefits (Gürbüz et al., 2018). Eggplant is rich in some key metabolites contributing to its functional and pharmaceutical properties (Rodriguez-Jimenez et al., 2018). Increased PAL and PPO activity (Table 4) under crop rotation with different vegetables justify the beneficial effect on secondary metabolism. The accumulation of secondary metabolites is regulated by PPO, which does this by oxidizing phenols. This, in turn, mediates fruit harvesting and resistance to pathogens. On the other hand, it has been reported that silencing PPO makes pathogen infection more likely by modifying the accumulation of phenolic compounds and their derivatives (Araji et al., 2014). However, through metabolomics increase in individual metabolites can be assessed to unravel the exact mechanisms involved.

5 Conclusion

Conclusively results of the study, which was carried out underneath a plastic shed using sustainable practices, indicated that different leafy vegetable species could be successfully used to minimize external inputs without a reduction in yield. The study was carried out within the context of the transition of agricultural practices toward the cultivation of sustainable vegetables. Conclusively, crop rotation of eggplant with leafy vegetable cropping systems, including WOE, CE, NCCE, and LE, exhibited greater yield and growth through improving soil chemical properties, modulation in the photosynthetic efficiency and gas exchange parameters. Increased activity of antioxidant enzymes imparted reduced oxidative damage by lowering the generation of reactive oxygen species. In addition, crop rotation with leafy vegetables may have regulated the metabolism of secondary metabolites through the upregulation of PAL and PPO. By modulating ROS and altering the activity of antioxidant enzymes, NCCE, and CE were more effective in improving growth and yield than other leafy vegetable species assessed in this study, including fallow eggplant. This was determined by comparing their results to

those of other leafy vegetable plants. Further studies at transcriptomic, metabolomic, and molecular levels would be helpful in unraveling the exact mechanisms of the above findings.

Data availability statement

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

Author contributions

ZC: Scientific concept, foundation, conceptualization, research idea, experimental guiding, revising, editing, and corresponding author, XC: review, foundation, editing, proof reading of the entire manuscript and co-corresponding author, MIG: Performing the field experiment, investigation, experimental guiding, data collection, methodology, and writing the draft manuscript, MAt, AL, MAl and MN: Investigation, accurateness of data analysis, and correction. MIG and MAh analyzed the data, interpreted the results, and wrote the paper with input from all authors. All authors contributed to the article and approved the submitted version.

Funding

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

Ahanger, M. A., and Agarwal, R. (2017a). Potassium up-regulates antioxidant metabolism and alleviates growth inhibition under water and osmotic stress in wheat (Triticum aestivum l). *Protoplasma* 254, 1471–1486. doi: 10.1007/s00709-016-1037-0

Ahanger, M. A., and Agarwal, R. (2017b). Salinity stress induced alterations in antioxidant metabolism and nitrogen assimilation in wheat (Triticum aestivum l) as influenced by potassium supplementation. *Plant Physiol. Biochem.* 115, 449–460. doi: 10.1016/j.plaphy.2017.04.017

Ahanger, M. A., Alyemeni, M. N., Wijaya, L., Alamri, S. A., Alam, P., Ashraf, M., et al. (2018). Potential of exogenously sourced kinetin in protecting solanum lycopersicum from NaCl-induced oxidative stress through up-regulation of the antioxidant system, ascorbate-glutathione cycle and glyoxalase system. *PloS One* 13, e0202175. doi: 10.1371/journal.pone.0202175

Ahanger, M. A., Mir, R. A., Alyemeni, M. N., and Ahmad, P. (2020). Combined effects of brassinosteroid and kinetin mitigates salinity stress in tomato through the modulation of antioxidant and osmolyte metabolism. *Plant Physiol. Biochem.* 147, 31–42. doi: 10.1016/j.plaphy.2019.12.007

Ahanger, M. A., Tomar, N. S., Tittal, M., Argal, S., and Agarwal, R. (2017). Plant growth under water/salt stress: ROS production; antioxidants and significance of added potassium under such conditions. *Physiol. Mol. Biol. Plants* 23, 731–744. doi: 10.1007/s12298-017-0462-7

Ahmad, P. (2010). Growth and antioxidant responses in mustard (Brassica juncea l.) plants subjected to combined effect of gibberellic acid and salinity. *Arch. Agron. Soil Sci.* 56, 575–588. doi: 10.1080/03650340903164231

Ali, A., Ghani, M. I., Elrys, A. S., Ding, H., Iqbal, M., Cheng, Z., et al. (2021). Different cropping systems regulate the metabolic capabilities and potential ecological functions altered by soil microbiome structure in the plastic shed mono-cropped cucumber rhizosphere. *Agriculture Ecosyst. Environ.* 318, 107486. doi: 10.1016/j.agee.2021.107486

Ali, A., Imran Ghani, M., Li, Y., Ding, H., Meng, H., and Cheng, Z. (2019). Hiseq base molecular characterization of soil microbial community, diversity structure, and predictive functional profiling in continuous cucumber planted soil affected by diverse cropping systems in an intensive greenhouse region of northern China. *Int. J. Mol. Sci.* 20, 2619. doi: 10.3390/ijms20112619

Araji, S., Grammer, T.A., Gertzen, R., Anderson, S. D., Mikulic-Petkovsek, M., Veberic, R., et al (2014). Novel roles for the polyphenol oxidase enzyme in secondary metabolism and the regulation of cell death in walnut. *Plant Physiol.* 164, 1191–1203. doi: 10.1104/pp.113.228593

Arnon, D. I. (1949). Copper enzymes in isolated chloroplasts. polyphenoloxidase in beta vulgaris. *Plant Physiol.* 24, 1. doi: 10.1104/pp.24.1.1

Aschi, A., Aubert, M., Riah-Anglet, W., Nélieu, S., Dubois, C., Akpa-Vinceslas, M., et al. (2017). Introduction of faba bean in crop rotation: Impacts on soil chemical and biological characteristics. *Appl. Soil Ecol.* 120, 219–228. doi: 10.1016/j.apsoil.2017.08.003

Bhatti, U. A., Yu, Z., Hasnain, A., Nawaz, S. A., Yuan, L., Wen, L., et al. (2022). Evaluating the impact of roads on the diversity pattern and density of trees to improve the conservation of species. *Environ. Sci. Pollut. Res.* 29, 14780–14790, 1-11. doi: 10.1007/s11356-021-16627-y

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Chahal, I., and Van Eerd, L. L. (2020). Cover crop and crop residue removal effects on temporal dynamics of soil carbon and nitrogen in a temperate, humid climate. *PloS One* 15, e0235665. doi: 10.1371/journal.pone.0235665

Chance, B., and Maehly, A. (1955). Assay of catalases and peroxidases. In SP Colowick and NO Kaplaneds. *Methods in Enzymology*. New York, NY: Academic Press. doi: 10.1016/S0076-6879(55)02300-8

Chen, D., Mubeen, B., Hasnain, A., Rizwan, M., Adrees, M., Naqvi, S. A. H., et al. (2022). Role of promising secondary metabolites to confer resistance against environmental stresses in crop plants: Current scenario and future perspectives. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.881032

Chen, S., Zhou, B., Lin, S., Li, X., and Ye, X. (2011). Accumulation of cinnamic acid and vanillin in eggplant root exudates and the relationship with continuous cropping obstacle. *Afr. J. Biotechnol.* 10, 2659–2665. doi: 10.5897/AJB10.1338

Cheng, F., and Cheng, Z. (2015). Research progress on the use of plant allelopathy in agriculture and the physiological and ecological mechanisms of allelopathy. *Front. Plant Sci.* 6, 1020. doi: 10.3389/fpls.2015.01020

Chuang, H.-W., Feng, J.-H., and Feng, Y.-L. (2016). The role of arabidopsis WDR protein in plant growth and defense strategies. *Plant Signaling Behav.* 11, 23–31. doi: 10.1080/15592324.2016.1217376

Crow, W., Weingartner, D., and Dickson, D. (2000). Effects of potato-cotton cropping systems and nematicides on plant-parasitic nematodes and crop yields. *J. Nematol.* 32, 297.

D'Acunto, L., Andrade, J. F., Poggio, S. L., and Semmartin, M. (2018). Diversifying crop rotation increased metabolic soil diversity and activity of the microbial community. *Agriculture Ecosyst. Environ.* 257, 159–164. doi: 10.1016/j.agee.2018.02.011

Dhindsa, R. S., Plumb-Dhindsa, P., and Thorpe, T. A. (1981). Leaf senescence: Correlated with increased levels of membrane permeability and lipid peroxidation, and decreased levels of superoxide dismutase and catalase. *J. Exp. Bot.* 32, 93–101. doi: 10.1093/jxb/32.1.93

Dorissant, L., Brym, Z. T., and Swartz, S. (2022). Residue decomposition dynamics in mixed two warm-season cover crops. *Agrosyst. Geosciences Environ.* 5, e20311. doi: 10.1002/agg2.20311

Gao, J. (2006). *Experimental guidance for plant physiology*. Beijing: Higher Education Press, 123–130.

Gaudin, A. C., Tolhurst, T. N., Ker, A. P., Janovicek, K., Tortora, C., Martin, R. C., et al. (2015). Increasing crop diversity mitigates weather variations and improves yield stability. *PloS One* 10, e0113261. doi: 10.1371/journal.pone.0113261

Ghani, M. I., Ali, A., Atif, M. J., Ali, M., Amin, B., Anees, M., et al. (2019a). Soil amendment with raw garlic stalk: A novel strategy to stimulate growth and the antioxidative defense system in monocropped eggplant in the north of China. *Agronomy* 9, 89. doi: 10.3390/agronomy9020089

Ghani, M. I., Ali, A., Atif, M. J., Ali, M., Amin, B., Anees, M., et al. (2019b). Changes in the soil microbiome in eggplant monoculture revealed by high-throughput illumina MiSeq sequencing as influenced by raw garlic stalk amendment. *Int. J. Mol. Sci.* 20, 2125. doi: 10.3390/ijms20092125

Ghani, M. I., Ali, A., Atif, M. J., Ali, M., Amin, B., and Cheng, Z. (2022a). Arbuscular mycorrhizal fungi and dry raw garlic stalk amendment alleviate continuous monocropping growth and photosynthetic declines in eggplant by bolstering its antioxidant system and accumulation of osmolytes and secondary metabolites. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.849521

Ghani, M. I., Ali, A., Atif, M. J., Pathan, S. I., Pietramellara, G., Ali, M., et al. (2022b). Diversified crop rotation improves continuous monocropping eggplant production by altering the soil microbial community and biochemical properties. *Plant Soil* 480, 603–624, 1–22. doi: 10.1007/s11104-022-05606-y

Ghani, M. I., Saleem, S., Rather, S. A., Rehmani, M. S., Alamri, S., Rajput, V. D., et al. (2022c). Foliar application of zinc oxide nanoparticles: An effective strategy to mitigate drought stress in cucumber seedling by modulating antioxidant defense system and osmolytes accumulation. *Chemosphere* 289, 133202. doi: 10.1016/j.chemosphere.2021.133202

Ghani, M. I., Wang, J., Li, P., Pathan, S. I., Sial, T. A., Datta, R., et al. (2022d). Variations of soil organic carbon fractions in response to conservative vegetation successions on the loess plateau of China. *Int. Soil Water Conserv. Res.* doi: 10.1016/j.iswcr.2022.05.002

Gürbüz, N., Uluişik, S., Frary, A., Frary, A., and Doğanlar, S. (2018). Health benefits and bioactive compounds of eggplant. *Food Chem.* 268, 602–610. doi: 10.1016/j.foodchem.2018.06.093

Jones, D. L., Nguyen, C., and Finlay, R. D. (2009). Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant Soil* 321, 5–33. doi: 10.1007/s11104-009-9925-0

Khan, M. I. R., and Khan, N. A. (2014). Ethylene reverses photosynthetic inhibition by nickel and zinc in mustard through changes in PS II activity, photosynthetic nitrogen use efficiency, and antioxidant metabolism. *Protoplasma* 251, 1007–1019. doi: 10.1007/s00709-014-0610-7

Knudsen, D., Peterson, G., and Pratt, P. (1983). Lithium, sodium, and potassium. *Methods Soil Analysis: Part 2 Chem. Microbiol. Properties* 9, 225–246.

Kong, A. Y., and Six, J. (2012). Microbial community assimilation of cover crop rhizodeposition within soil microenvironments in alternative and conventional cropping systems. *Plant Soil* 356, 315–330. doi: 10.1007/s11104-011-1120-4

Lee, M. J., and Yaffe, M. B. (2016). Protein regulation in signal transduction. *Cold Spring Harbor Perspect. Biol.* 8, a005918. doi: 10.1101/cshperspect.a005918

Li, T., Liu, T., Zheng, C., Kang, C., Yang, Z., Yao, X., et al. (2017). Changes in soil bacterial community structure as a result of incorporation of brassica plants compared with continuous planting eggplant and chemical disinfection in greenhouses. *PloS One* 12, e0173923. doi: 10.1371/journal.pone.0173923

Liu, Z., Gao, F., Yang, J., Zhen, X., Li, Y., Zhao, J., et al. (2019). Photosynthetic characteristics and uptake and translocation of nitrogen in peanut in a wheat–peanut rotation system under different fertilizer management regimes. *Front. Plant Sci.* 10, 86. doi: 10.3389/fpls.2019.00086

Long, A., Zhang, J., Yang, L.-T., Ye, X., Lai, N.-W., Tan, L.-L., et al. (2017). Effects of low pH on photosynthesis, related physiological parameters, and nutrient profiles of citrus. *Front. Plant Sci.* 8, 185. doi: 10.3389/fpls.2017.00185

Lyu, J., Jin, L., Jin, N., Xie, J., Xiao, X., Hu, L., et al. (2020). Effects of different vegetable rotations on fungal community structure in continuous tomato cropping matrix in greenhouse. *Front. Microbiol.* 11, 829. doi: 10.3389/fmicb.2020.00829

Murphy, J., and Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Analytica chimica Acta* 27, 31–36. doi: 10.1016/S0003-2670(00)88444-5

Nahar, K., Hasanuzzaman, M., Alam, M. M., Rahman, A., Suzuki, T., and Fujita, M. (2016). Polyamine and nitric oxide crosstalk: Antagonistic effects on cadmium toxicity in mung bean plants through upregulating the metal detoxification, antioxidant defense and methylglyoxal detoxification systems. *Ecotoxicol. Environ. Saf.* 126, 245–255. doi: 10.1016/j.ecoenv.2015.12.026

Nakano, Y., and Asada, K. (1981). Hydrogen peroxide is scavenged by ascorbate-specific peroxidase in spinach chloroplasts. *Plant Cell Physiol.* 22, 867–880. doi: 10.1093/oxfordjournals.pcp.a076232

Radicetti, E., Massantini, R., Campiglia, E., Mancinelli, R., Ferri, S., and Moscetti, R. (2016). Yield and quality of eggplant (Solanum melongena l.) as affected by cover crop species and residue management. *Scientia Hortic.* 204, 161–171. doi: 10.1016/j.scienta.2016.04.005

Rodriguez-Jimenez, J. R., Amaya-Guerra, C. A., Baez-Gonzalez, J. G., Aguilera-Gonzalez, C., Urias-Orona, V., and Nino-Medina, G. (2018). Physicochemical, functional, and nutraceutical properties of eggplant flours obtained by different drying methods. *Molecules* 23, 3210. doi: 10.3390/molecules23123210

Sánchez-Navarro, V., Zornoza, R., Faz, Á., and Fernández, J. A. (2020). Comparison of soil organic carbon pools, microbial activity and crop yield and quality in two vegetable multiple cropping systems under mediterranean conditions. *Scientia Hortic.* 261, 109025. doi: 10.1016/j.scienta.2019.109025

Scarascia-Mugnozza, G., Sica, C., and Russo, G. (2011). Plastic materials in European agriculture: Actual use and perspectives. *J. Agric. Eng.* 42, 15–28. doi: 10.4081/jae.2011.28

Shi, R. (1996). *Agricultural Chemistry Analyses of Soils*, *2nd Edn. Beijing*: China Agricultural Press (in Chinese), 37–39.

Song, X., Tao, B., Guo, J., Li, J., and Chen, G. (2018). Changes in the microbial community structure and soil chemical properties of vertisols under different cropping systems in northern China. *Front. Environ. Sci.* 6, 132. doi: 10.3389/fenvs.2018.00132

St. Luce, M., Lemke, R., Gan, Y., Mcconkey, B., May, W., Campbell, C., et al. (2020). Diversifying cropping systems enhances productivity, stability, and nitrogen use efficiency. *Agron. J.* 112, 1517–1536. doi: 10.1002/agj2.20162

Su, B., Song, Y., Song, C., Cui, L., Yong, T., and Yang, W. (2014). Growth and photosynthetic responses of soybean seedlings to maize shading in relay intercropping system in southwest China. *Photosynthetica* 52, 332–340. doi: 10.1007/s11099-014-0036-7

Tan, G., Liu, Y., Peng, S., Yin, H., Meng, D., Tao, J., et al. (2021). Soil potentials to resist continuous cropping obstacle: Three field cases. *Environ. Res.* 200, 111319. doi: 10.1016/j.envres.2021.111319

Thorup-Kristensen, K., Dresbøll, D. B., and Kristensen, H. L. (2012). Crop yield, root growth, and nutrient dynamics in a conventional and three organic cropping systems

with different levels of external inputs and n re-cycling through fertility building crops. *Eur. J. Agron.* 37, 66–82. doi: 10.1016/j.eja.2011.11.004

Tian, Y., Zhang, X., Liu, J., Chen, Q., and Gao, L. (2009). Microbial properties of rhizosphere soils as affected by rotation, grafting, and soil sterilization in intensive vegetable production systems. *Scientia Hortic.* 123, 139–147. doi: 10.1016/j.scienta.2009.08.010

Tian, Y., Zhang, X., Wang, J., and Gao, L. (2013). Soil microbial communities associated with the rhizosphere of cucumber under different summer cover crops and residue management: A 4-year field experiment. *Scientia Hortic.* 150, 100–109. doi: 10.1016/j.scienta.2012.10.025

Velikova, V., Yordanov, I., and Edreva, A. (2000). Oxidative stress and some antioxidant systems in acid rain-treated bean plants: Protective role of exogenous polyamines. *Plant Sci.* 151, 59–66. doi: 10.1016/S0168-9452(99)00197-1

Walkley, A., and Black, I. A. (1934). An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* 37, 29–38. doi: 10.1097/00010694-193401000-00003

Wang, M., Wu, C., Cheng, Z., and Meng, H. (2015). Growth and physiological changes in continuously cropped eggplant (Solanum melongena l.) upon relay intercropping with garlic (Allium sativum l.). *Front. Plant Sci.* 6, 262. doi: 10.3389/fpls.2015.00262

Wu, F.-Z., and Wang, X.-Z. (2007). "*Effect of monocropping and rotation on soil microbial community diversity and cucumber yield and quality under protected cultivation*," (International Society for Horticultural Science (ISHS), Leuven, Belgium), 555–561.

Xiao, X., Cheng, Z., Meng, H., Khan, M. A., and Li, H. (2012). Intercropping with garlic alleviated continuous cropping obstacle of cucumber in plastic tunnel. *Acta Agriculturae Scandinavica Section B–Soil Plant Sci.* 62, 696–705. doi: 10.1080/09064710.2012.697571

Yao, X., Zhou, H., Zhu, Q., Li, C., Zhang, H., Wu, J.-J., et al. (2017). Photosynthetic response of soybean leaf to wide light-fluctuation in maize-soybean intercropping system. *Front. Plant Sci.* 8, 1695. doi: 10.3389/fpls.2017.01695

Yin, X. M., Huang, L. F., Zhang, X., Wang, M. L., Xu, G. Y., and Xia, X. J. (2015). OsCML4 improves drought tolerance through scavenging of reactive oxygen species in rice. *J. Plant Biol.* 58, 68–73. doi: 10.1007/s12374-014-0349-x

Zeng, J., Liu, J., Lu, C., Ou, X., Luo, K., Li, C., et al. (2020). Intercropping with turmeric or ginger reduce the continuous cropping obstacles that affect pogostemon cablin (patchouli). *Front. Microbiol.* 11, 579719. doi: 10.3389/fmicb.2020.579719

Zheng, L., Zhong, G., and Zhang, S. (2007). PPO activity determination and anti-browning measure for amorphophallus albus. *Chin. J. Southwest Univ. (Natural Sci. Edition)* 29, 118–121.

Check for updates

A YOLOv7 incorporating the Adan optimizer based corn pests identification method

Chong Zhang[1], Zhuhua Hu[1]*, Lewei Xu[1] and Yaochi Zhao[2]

[1]School of Information and Communication Engineering, State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China, [2]School of Cyberspace Security, State Key Laboratory of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China

Major pests of corn insects include corn borer, armyworm, bollworm, aphid, and corn leaf mites. Timely and accurate detection of these pests is crucial for effective pests control and scientific decision making. However, existing methods for identification based on traditional machine learning and neural networks are limited by high model training costs and low recognition accuracy. To address these problems, we proposed a YOLOv7 maize pests identification method incorporating the Adan optimizer. First, we selected three major corn pests, corn borer, armyworm and bollworm as research objects. Then, we collected and constructed a corn pests dataset by using data augmentation to address the problem of scarce corn pests data. Second, we chose the YOLOv7 network as the detection model, and we proposed to replace the original optimizer of YOLOv7 with the Adan optimizer for its high computational cost. The Adan optimizer can efficiently sense the surrounding gradient information in advance, allowing the model to escape sharp local minima. Thus, the robustness and accuracy of the model can be improved while significantly reducing the computing power. Finally, we did ablation experiments and compared the experiments with traditional methods and other common object detection networks. Theoretical analysis and experimental result show that the model incorporating with Adan optimizer only requires 1/2-2/3 of the computing power of the original network to obtain performance beyond that of the original network. The mAP@[.5:.95] (mean Average Precision) of the improved network reaches 96.69% and the precision reaches 99.95%. Meanwhile, the mAP@[.5:.95] was improved by 2.79%-11.83% compared to the original YOLOv7 and 41.98%-60.61% compared to other common object detection models. In complex natural scenes, our proposed method is not only time-efficient and has higher recognition accuracy, reaching the level of SOTA.

1 Introduction

In the past decade, due to the excellent performance of machine learning and deep learning techniques on other tasks, scholars have applied them to crop pests and disease identification and have made good, progress in pests and disease identification. Scholars have applied them to crop pests and disease identification and made good progress. In 2010, Al-Bashish et al. (Al Bashish et al., 2010a). Introduced proposed the use of K-means clustering with HSI color space co-occurrence to extract color and texture features of plants, ultimately classifying five different plant diseases with a simple neural network. Since then, research works based on various machine learning methods to identify plant diseases and pests have emerged. In 2016, Sladojevic et al. (Al Bashish et al., 2010a). developed a new method for identifying 13 different plant diseases using deep convolutional neural networks, achieving a final accuracy of 96.3%. The authors created a comprehensive database and methodology for modeling, which is essential for future research in this field. Scholars have gradually realized the great potential of deep learning techniques, and research on pests and disease identification based on various deep learning methods has proliferated. For instance, Amara et al. (2017) identified and classified banana leaf diseases in the natural environment by using LeNet network, Nachtigall et al. (Bochkovskiy et al., 2020). used CNN to recognize diseases, nutrient deficiencies and herbicide damage in apple leaf images. Inspired by these previous works, our team conducted research on corn borer and anthracnose spore identification using different machine learning methods, all of which yielded promising result. However, these traditional machine learning and deep learning methods above still have their own limitations, such as high model training cost and poor robustness, which sharply increase the cost of academic research or industrial implementation. Therefore, it is important to find a pests identification method with low training cost, accurate identification and good robustness.

1.1 Related work and motivation

With the development of digital image processing and machine learning techniques, intelligent detection and identification of crop diseases and pests have become increasingly prevalent. In plant disease identification, Sasaki et al. (Girshick, 2015). utilized spectral reflectance differences to distinguish healthy and diseased areas on cucumber leaves, while Vízhányó et al. (Girshick et al., 2014). used color point differences to identify diseased mushrooms. In China, Guili Xu et al. (He et al., 2017). achieved over 70% accuracy in identifying tomato leaves based on histogram-based color feature extraction. Yuxia Zhao et al. (Li et al., 2022). used a Bayesian classifier to successfully identify five diseases, including maize rust. Our team has proposed several algorithms, such as the marker watershed algorithm (Lin et al., 2017a) and the Otsu separation and symbolic similarity-driven level set algorithm (Lin et al., 2017b), for accurate statistics of anthracnose spore distribution density on farms for better control. Additionally, our team proposed an accurate segmentation method for diseased fruits based on log

similarity-constrained Otsu and distance rule level set activity profile evolution (Liu et al., 2019), which can achieve good segmentation of diseased fruits.

In the field of plant pests identification, various methods have been proposed to improve the accuracy and efficiency of the identification process. However, most of these methods have limitations that need to be addressed. For instance, In terms of plant pests identification, Prof. Zorui Shen of China Agricultural University (Liu et al., 2008) firstly used mathematical morphology to solve the problem and achieved good result, but the variation of the selection of structural elements in mathematical morphology will affect the identification result, then it will cause the robustness of the identification algorithm is not strong. For insects' color characteristics, Dr. Zhu used color histogram and double-tree complex wavelet transform (Liu et al., 2016) and support vector machine (Mohanty et al., 2016) to further improve the recognition rate, but this method requires reliable data sets for training, so a large number of images need to be acquired and the cost is high. In addition, Dr. Zhu also proposed the color histogram combined with Weber descriptors for insect recognition of Lepidoptera (Nachtigall et al., 2016), CART-based combined with LLC (Redmon et al., 2016), and color-based combined with OpponentSIFT features (Redmon and Farhadi, 2017). However, these methods require manual extraction of features and are not applicable to borer moth family pests. To address these limitations, we propose an automatic pests monitoring robotic vehicle with a Pyralidae recognition scheme based on histogram and multi-template image reverse mapping method (Redmon and Farhadi, 2018). This new approach enables the automatic capture of pests images and achieves a recognition accuracy of up to 94.3% in the natural farm planting scenario. We also propose a pests image segmentation method based on GMM and DRLSE (Ren et al., 2015), which can automatically identify positive and negative samples of specific pests from a large number of scene images with recognition accuracy of up to 95%. Additionally, our proposed hybrid Gaussian model-based texture disparity representation and texture disparity-guided DRLSE model (Sammany and Medhat, 2007) can also achieve accurate segmentation of crop pests and diseases.

While the traditional machine learning methods have contributed to the field of crop pests and disease identification, they have certain limitations that prevent them from achieving the desired result. The advancements in deep learning technology have paved the way for researchers to apply deep learning algorithms to pests recognition, resulting in significant progress in this field. Deep learning algorithms can automatically extract image features, making good use of this information to achieve high accuracy in pests and disease identification. Several studies have used deep learning techniques to identify and classify pests, achieving higher robustness, generalization, and accuracy. For example, Sammany et al. (Sasaki et al., 1998). utilized genetic algorithms to improve neural networks, reducing the dimensionality of feature vectors and improving pests recognition efficiency. Similarly, Al Bashish et al. (Sladojevic et al., 2016). used the K-means clustering algorithm to classify images into clusters, extracted feature values of color and texture for each cluster, and inputted them into neural networks for

classification. Mohanty et al. (Tian et al., 2019). used the GoogleNet convolutional neural network structure to build a pests identification model with satisfactory result. Compared to traditional machine learning methods, deep neural network-based pests recognition methods have better accuracy, making them an important research direction in pests recognition. As deep learning technology continues to advance, we can expect more breakthroughs in crop pests and disease identification, which will undoubtedly benefit the agriculture industry.

Deep learning models have shown promising result in identifying and detecting pests. However, there are still limitations that need to be addressed. In recent years, various sophisticated training methods have been developed to improve the generalization and robustness of deep models. Nevertheless, the cost of training these models has increased significantly due to the higher computing power required. This increase in training cost has a considerable impact on the research and industrial implementations. One common approach to reduce the training time is to increase the batch size and assist parallel training. However, a larger batch size often leads to a decrease in performance. The YOLOv7 method (Vızhányó and Felföldi, 2000), which is the current SOTA in object detection, also faces the same challenge. In this context, a new YOLOv7 corn pests identification method is proposed in this paper, which incorporates the Adan optimizer. This new method uses Adan (Wang et al., 2022), a novel optimizer that can sense the surrounding gradient information and efficiently escape from sharp local minimal areas. By replacing the original optimizer of YOLOv7 with Adan, the model can achieve faster and better training without compromising its accuracy. The proposed YOLOv7 method can identify major corn pests in complex natural environments quickly and accurately, reducing the cost of practical application of model. With fewer parameter updates, the deep model can achieve faster and more accurate identification, making it suitable for various applications. In summary, the YOLOv7 corn pests identification method incorporating the Adan optimizer presented in this paper can significantly reduce the training time and cost while maintaining the accuracy of the model. It is expected to contribute to the efficient and accurate identification of pests in agricultural production.

1.2 Contributions

1. To address the lack of maize pests data, we used data augmentation techniques to construct a maize pests image dataset, which effectively improved the training of the model.
2. We replaced the original optimizer of YOLOv7 with a new optimizer, Adan, which combines a rewritten Nesterov momentum algorithm with an adaptive optimization algorithm and introduces decoupled weight decay, allowing the model to increase its speed without degrading its accuracy, thus enabling faster and better training of the model and reducing the cost of implementing the model.

3. From the theoretical analysis and experimental result, it can be seen that the YOLOv7 network incorporating the Adan optimizer can effectively alleviate the negative impact caused by the increase of batch size, and solve the problem that the training speed and training accuracy cannot be achieved at the same time.

1.3 Paper organization

The rest of this paper is organized as follows. The second part Section 2 mainly introduces the related network model; In the third part Section 3, experimental scheme, process and results are introduced in detail; The fourth part Section 4 discusses the experimental results; The fifth part Section 5 summarizes the full text and puts forward the existing deficiencies and the direction that can be improved.

2 Materials and methods

This section first introduces the basic concepts of object detection network. Then it describes the YOLOv7 network and Adan optimizer used in this project, and finally introduces the proposed improved network.

2.1 Object detection network

Object detection is one of the core problems in the field of computer vision. It needs to find out all the objects of interest in an image, and determine their classes and locations. Object detection is always the most challenging problem in the field of computer vision because of the different appearances, shapes and poses of various objects, as well as the interference of illumination, occlusion and other factors during imaging. A diagram of the object detection task is shown in Figure 1.

The current popular algorithms can be divided into two categories, one is the two-stage algorithm based on Region Proposal, which find out some candidate regions primarily, and then adjust the regions for classification, such as the series of R-CNN (Regions with CNN features) algorithm (Xu et al., 2002; Zhao and Hu, 2015; Wang et al., 2020; Xie et al., 2022). The other category is one-stage algorithm, such as SSD (Zhao et al., 2015) (Single Shot Multibox Detector), the series of YOLO (You Only Look Once) algorithm (Vızhányó and Felföldi, 2000; Zhao et al., 2007; Zhu et al., 2015a; Zhu et al., 2015b; Zhao et al., 2019; Zhao et al., 2020), RetinaNet (Zhu et al., 2010), FCOS (Zhu et al., 2012) (Fully Convolutional One-Stage Object Detection) and other such side-to-side networks. They only use a convolutional neural network to directly predict classes and locations of different objects. Comparing the two categories of object detection algorithms, the former is more accurate but slower, while the latter is faster but less accurate. In this paper, some representative
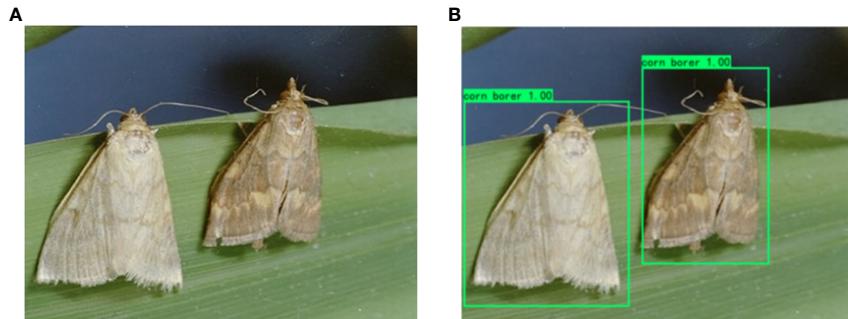
FIGURE 1
Schematic diagram of object detection: **(A)** Original map. **(B)** Object detection map.

networks in the above two categories are selected for comparative experiments.

2.2 YOLOv7

YOLOv7 is a new network framework based on the series of YOLO algorithm, which mainly designs a better performance detection model through the following four aspects: backbone design with new ELAN module, composite model scaling, deep supervision label assignment strategy and model re-parameterization.

The first improvement is the design of new network structure. YOLOv7 proposes such a view: the shortest and longest gradient paths can be controlled to achieve more effective learning and convergence of deep networks. Based on this idea, YOLOv7 designs the E-ELAN network structure as shown in Figure 2 on the basis of ELAN. In common ELAN module, the whole network reaches a stable state regardless of the gradient path length and the number of computing modules. However, if more ELAN modules are stacked indefinitely, this stable state may be destroyed and the parameter utilization may be reduced. Based on the above shortcomings, YOLOv7 proposes the E-ELAN module. E-ELAN module adopts the structure of expand, shuffle and merge cardinality, and it can guide different computing blocks to learn more diversified characteristics compared to common ELAN module, thus improving the learning ability of the network without destroying the original gradient path.

The second improvement is composite model scaling. The main purpose of model scaling is to adjust certain properties of the model and generate models of different sizes to meet the needs of different inference speeds. If the E-ELAN method described above is applied directly to a cascaded model, the action of directly scaling up the depth of the model will result in a change in the scale of the input and output channels. As a result, the model's usage of hardware may decrease. Therefore, for the cascaded model, a composite model method must be proposed. The method must consider that the width of the transition layer should also be changed by the same amount when the depth of the computing module is scaled. Based on these ideas, YOLOv7 proposes a network architecture as shown in Figure 3. The network only needs to scale the depth in the

computation block when performing the model scaling, and the rest of the transport block will use the corresponding width scaling. The composite scaling method can preserve the properties of the model at the initial design and maintain the optimal structure.

The third improvement is deep supervision label assignment strategy. Deep supervision is a common technique in deep network training, it adds auxiliary head for loss calculation in the middle of the network to assist training. In order to differentiate auxiliary head for different functions, the final output head is called the Lead Head and the auxiliary training head is called the Aux Head. The core idea of deep supervision is to take shallow network weight and auxiliary loss as guidance, combine the output result with Ground
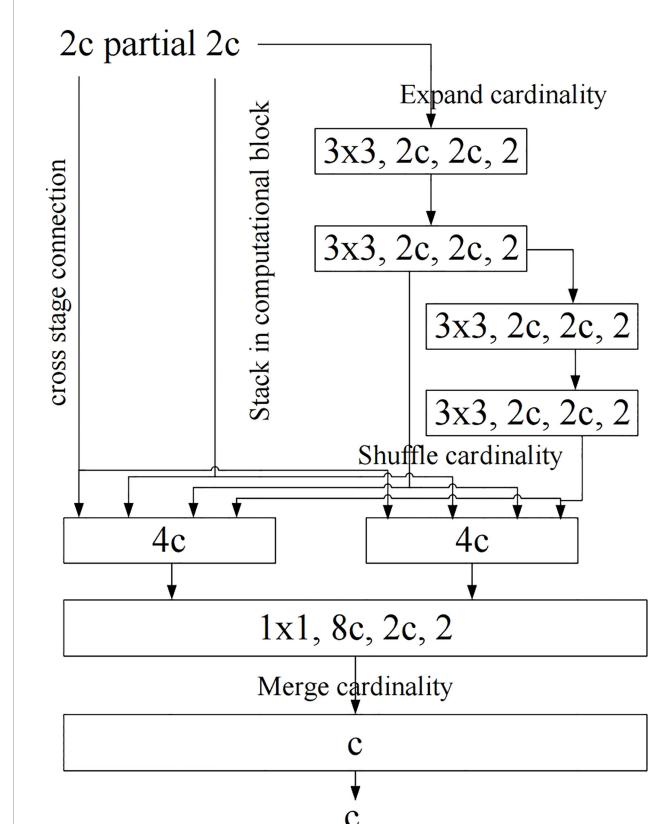

FIGURE 2
E-ELAN structure diagram. bold values means the better results.
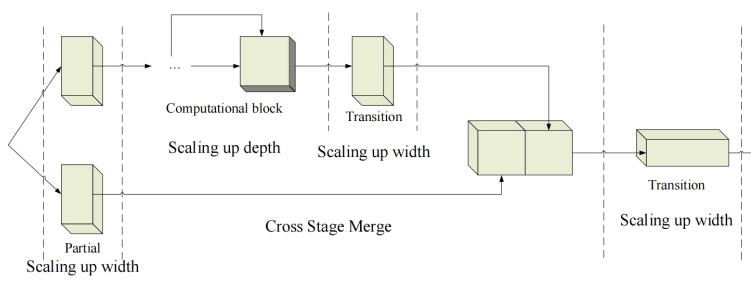
FIGURE 3
Composite model scaling for YOLOv7. bold values means the better results.

True (GT), and use some calculation and optimization methods to generate reliable soft labels. For example, YOLO uses the bounding box regression and GT and the IOU of the prediction box as soft labels. The current common method of assigning soft labels to Aux Head and Lead Head is shown in the Figure 4A, which separates Aux Head and Lead Head, and uses their respective prediction result and GT to perform label assignment. In contrast, YOLOv7 network uses the Lead Head prediction result as a guide to generate coarse-to-fine hierarchical labels for Aux Heads and other Lead Heads learning. The two proposed deep supervision label assignment strategies are shown in Figures 4B, C. The reason for this is that the Lead Head has strong learning ability, and the generated soft labels should better represent the distribution and correlation between the source and target data. By allowing the shallow Aux Heads to directly learn the information that Lead Heads has already learned, the Lead Heads will be better able to focus on learning residual information that has not yet been learned.

The last improvement is model re-parameterization. Re-parameterization is a technique used to improve a model after training, which increases the training time but improves the inference result. Although model re-parameterization has achieved excellent performance on VGG, when applied directly to architectures such as ResNet and DenseNet, it instead causes a significant decrease in accuracy. For these reasons, YOLOv7 uses the constant connection-free RepConvN to redesign the architecture of the reparameterized convolution by replacing the

3×3 convolutional layers of the E-ELAN computational block with constant connection-free RepConv layers.

2.3 Adan optimizer

The most direct way to speed up the convergence of the optimizer is to import momentums. The deep model optimizers proposed in recent years all follow the same momentum paradigm used in Adam - the reball method. However, with the advent of ViT, researchers found that Adam was not able to train ViT. And AdamW, an improved version of Adam, gradually became the preferred choice for training ViT and even ConvNext. However, AdamW does not change the momentum paradigm in Adam, which tends to cause the performance of AdamW-trained networks to drop dramatically when the batch size increases to a certain threshold.

In the field of traditional convex optimization, there is an momentum algorithm equal to the heavy ball method, the Nesterov momentum algorithm. As shown in Equation 1.

$$AGD : g_k = \nabla f(\theta_k - \eta(1 - \beta_1)m_{k-1}) + \xi_k, m_k$$
$$= (1 - \beta_1)m_{k-1} + g_k, \theta_{k+1} = \theta_k - \eta m_k \tag{1}$$

The Nesterov momentum algorithm has a faster theoretical convergence rate than the heavy ball method for smooth and generally convex problems, and can theoretically withstand larger
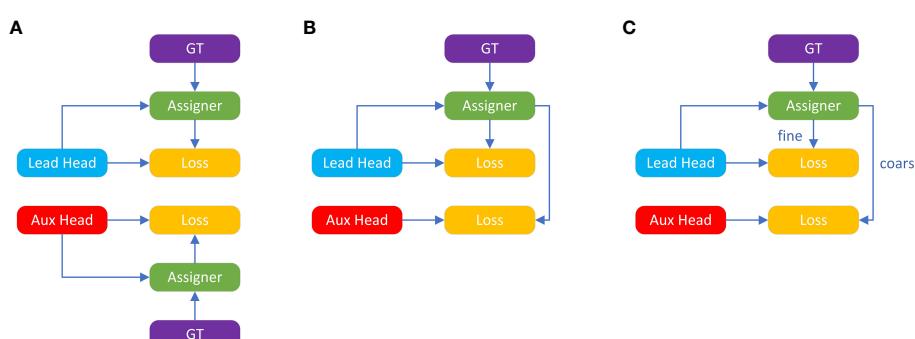

FIGURE 4
Deep supervision label assignment strategies: **(A)** Common strategy. **(B, C)** Two proposed strategies of YOLOv7. bold values means the better results.

batch size. Different from the heavy ball method, Nesterov algorithm does not calculate the gradient at the current point, but uses the momentum to find an extrapolation point, and then carries on the momentum accumulation after calculating the gradient at the point. Although Nesterov momentum algorithm has some advantages, it is rarely applied and explored in depth optimizers. One of the main reasons is that Nesterov algorithm needs to calculate gradient at extrapolated points, which requires multiple overloading of model parameters during updating at current points and requires artificial back-propagation (BP) at extrapolated points. These inconveniences greatly limit the application of Nesterov momentum algorithm in depth model optimizer.

In order to give full play to the advantages of the Nesterov momentum algorithm, Adan researchers obtained the final Adan optimizer by combining the rewritten Nesterov momentum with the adaptive optimization algorithm and introducing decoupled weight attenuation. In order to solve the problem of multiple model parameter overloads in the Nesterov momentum algorithm, the researchers first rewrote the Nesterov momentum algorithm as shown in Equation 2.

$$\text{Reformulated AGD}: \begin{cases} g_k = E_{\zeta \sim D}[\nabla f(\theta_k, \zeta)] + \xi_k \\ m_k = (1 - \beta_1)m_{k-1} + [g_k + (1 - \beta_1)(g_k - g_{k-1})] \\ \theta_{k+1} = \theta_k - \eta m_k \end{cases}$$

$$(2)$$

Combining the rewritten Nesterov momentum algorithm with the adaptive class optimizer - replacing the update of m_k from the cumulative form to the moving average form and using the second-order moment to deflate the learning rate - has resulted in a basic version of Adan's algorithm. As shown in Equation 3.

$$\text{Vanilla Adan}: \begin{cases} m_k = (1 - \beta_1)m_{k-1} + \beta_1[g_k + (1 - \beta_1)(g_k - g_{k-1})] \\ n_k = (1 - \beta_3)n_{k-1} + \beta_3[g_k + (1 - \beta_1)(g_k - g_{k-1})]^2 \\ n_k = \eta / \sqrt{n_k + \epsilon} \\ \theta_{k+1} = \theta_k - \eta_k \circ m_k \end{cases}$$

$$(3)$$

Although it can be seen that the update of m_k combines the gradient with the gradient's difference, in real-world applications it is frequently necessary to treat the two physically distinct meaningful things separately. For this reason, the researchers developed the gradient difference momentum v_k, as shown in Equation 4.

$$m_k = (1 - \beta_1)m_{k-1} + \beta_1 g_k, \quad v_k = (1 - \beta_2)v_{k-1} + \beta_2(g_k - g_{k-1}) \quad (4)$$

Here different momentum/average coefficients are set for the momentum of the gradient and its difference. The gradient difference term can slow down the optimizer update when adjacent gradients are not consistent and, conversely, speed up the update when the gradients are in the same direction.

Based on the idea of L2 regular decoupling, Adan introduces a weight attenuation strategy, each iteration of Adan can be regarded as minimizing some first-order approximation of the optimization objective F, as shown in Equation 5.

$$\theta_{k+1} = \theta_k - \eta_k \circ \bar{m}_k = \underset{\theta}{\arg\min} \left(F(\theta_k) + \langle \bar{m}_k, \beta - \theta_k \rangle + \frac{1}{2\eta} ||\theta - \theta_k||^2_{\sqrt{n_k}} \right),$$

$$where \quad ||x||^2_{\sqrt{n_k}} := \langle x, \sqrt{n_k + \epsilon} \circ x \rangle, \quad \bar{m}_k := m_k + (1 - \beta_2)v_k$$

$$(5)$$

Because L2 weight regularization in F is too simple and smooth, it is unnecessary to make a first-order approximation. Therefore, only the first-order approximation of training loss can be performed and L2 weight regularization can be ignored. Then the last iteration of Adan will become as shown in Equation 6.

$$\theta_{k+1} = \theta_k - \eta_k \circ \bar{m}_k = \underset{\theta}{\arg\min} F(\theta_k) + \bar{m}_k, \theta - l\theta_k + \frac{1}{2\eta} ||\theta - \theta_k||^2_{\sqrt{n_k}}$$

$$(6)$$

The final Adan optimization algorithm can be obtained by combining the above two improvements Equation 4 and Equation 6 into the base version of Adan.

2.4 The proposed identification method

Since the network architecture is not changed, we still use the original network structure of YOLOv7, as shown in Figure 5.

After replacing the optimizer inside YOLOv7 with Adan, the loss function module will calculate the loss of this forward inference according to the difference between the output of model and the real label. Subsequently, the model will take the derivative of loss to obtain the gradient of each learnable parameter. Then the Adan optimizer can obtain the gradient and update parameters through the optimization strategy described above, such as m_k, v_k, n_k, etc. The model keeps the loss decreasing by updating these parameters after each inference, thus gradually reducing the difference between the output of model and the real label, and finally achieving the convergence. The whole model training process is shown in Figure 6, and the pseudocode is shown in algorithm 1.

```
Input: An image [H×W×3].
Output: Detection image.
Preprocessing: The input RGB image aligned to
an RGB image of size 640×640.
Training
for every image in training set do
    Stage 1: The processed images are input into
        the backbone module for feature
        extraction, while the backbone module will
        output three feature maps in different
        scales. And these feature maps will be
        input into the head module together for
        prediction.
    Stage 2: In the head module, three types of
        feature maps will be fused and input into
        RepVGG block and detect block to predict
        objects.
    Stage 3: The loss function module will
        calculate the loss of this inference
```
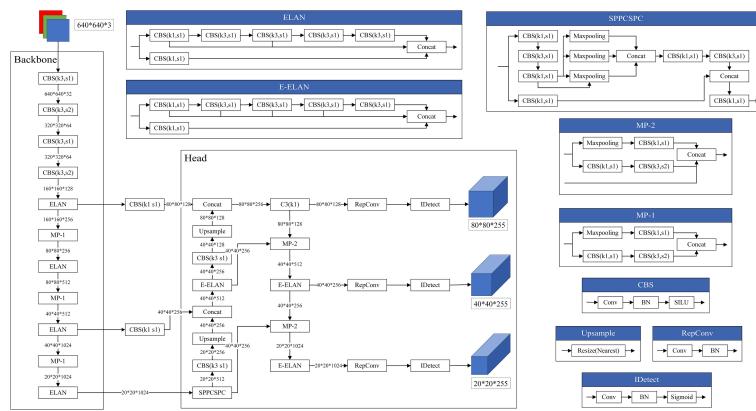
FIGURE 5
Network structure diagram.

according to the difference between the output of model and the real label. Subsequently, the model will take the derivative of loss to get the gradient and pass it to the optimizer module.

Stage 4: Adan Optimizer will initialize the following parameters: initialization θ_0, step size η, average parameter $(\beta_1, \beta_2, \beta_3, \in [0,1]^3)$, stable parameter, weight decays $\epsilon > 0$ and restart condition $\lambda_k > 0$, and then start the optimizing strategy.

for $k < K$ **do**

compute the stochastic gradient estimator g_k at θ_k;

$mk = (1 - \beta1)\, mk - 1 + \beta1 gk /*set\ m0 = g0*/$

$vk = (1 - \beta2)\, vk - 1 + \beta2(gk - gk - 1)/*set\ v1 = g1 - g0*/;$

$nk = (1 - \beta3)\, nk - 1 + \beta3[gk + (1 - \beta2)(gk - gk - 1)]^2$

$nk = \eta/(\sqrt{nk + \epsilon})$

if *restart condition holds* **then**
$\theta k + 1 = (1 + \lambda k \eta) - 1[\theta k - \eta k^{-1} \circ (mk + (1 - \beta2)vk)]$

get stochastic gradient estimator g_0 at θ_{k+1};

$m_0 = g_0,\ v_{0=0},\ n_0 = (g_0)^2$, update θ_1 by Line 7, $k = 1$;

```
        end
      end
end
```

ALGORITHM 1
Description of the algorithm of YOLOv7 incorporating the Adan optimizer

3 Experiments and result

3.1 Experimental scheme

The experimental scheme proposed is shown as Figure 7. We first pre-processed the original dataset, mainly including data recovery, data filtering and data filling. In order to solve the problem of scarce data, we used data augmentation and transfer learning to ensure that the network can fully learn the features. The two technologies will be introduced in detail in the following sections. And then, the augmented dataset was divided into training set, testing set and validation set. The training set and validation set was input into the original YOLOv7 network, the improved YOLOv7 network and other comparative networks respectively. If the performance of the model does not meet expectations, we will adjust the network's hyperparameters and retrain it. After that, the testing set was input into trained models to test the performance of different models. Finally, we compared and analyzed the experimental result.
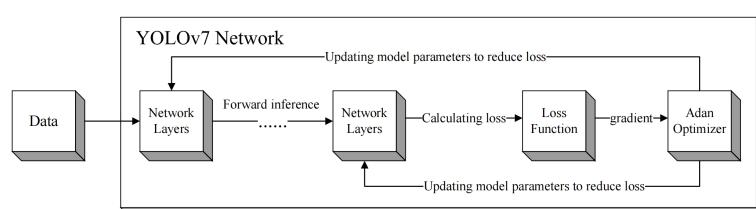

FIGURE 6
Flow chart of model training.
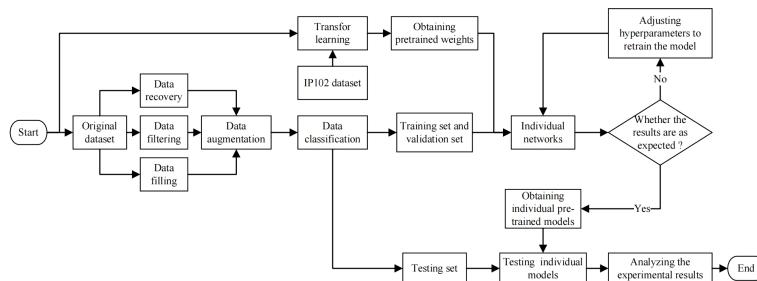
FIGURE 7
Flow chart of experimental scheme.

3.2 Evaluation metrics

For binary classification problem, A is called "positive" and B is called "negative," and the classifier correctly predicts "True" and incorrectly predicts "False". According to these four basic combinations, the four basic elements of the confusion matrix are TP (True Positive), FN (False Negative), TN ((True Negative), and FP (False Positive), as shown in Table 1.

In object detection experiments, IoU, Precision, Recall, AP and mAP are commonly used as evaluation indexes. Among them, IoU represents the intersection ratio between the predicted result and the true label for each category, as shown in Eq.7. Precision refers to the proportion of data whose value is true indeed when the classifier predicts it to be true, while Recall refers to the percentage of the classifier predicts to be correct for all data that is true, respectively, the formulas of the two is Eq.8 and Eq.9. However, all three indexes have their limitations, therefore AP/mAP is often used to evaluate the performance of object detection task.

$$IoU = \frac{TP}{TP+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

If we take different confidence levels, we can get different Precision and Recall, and if we get the confidence level dense enough, we will obtain the Precision-Recall curve(PR curve), as shown in Figure 8. While AP refers to the area under the curve, and

mAP is the average of the AP values for all classes. In particular, the mAP@[.5:.95] refers to the mAP at different IoU thresholds (from 0.5 to 0.95, in steps of 0.05).

3.3 Dataset acquisition

Due to the scarcity of public corn pests dataset, we collected some images of three major corn pests: corn borer, bollworm, and armyworm on the web as our original dataset, including 31 images of corn borer, 36 images of bollworm, 31 images of armyworm and 31 negative images. Prior to beginning the experiment, we used data augmentation techniques to the technique expands a total of 129 images to 5160 images as our final dataset. During training, we use a ratio of 8:1:1 to split the dataset into a training set, a validation set and a testing set. And the training set has 4128 images, the validation set has 516 images and the testing set has 516 images.

TABLE 1 Confusion matrix.

Truth	Prediction	
	T	F
P	TP	FN
N	FP	TN
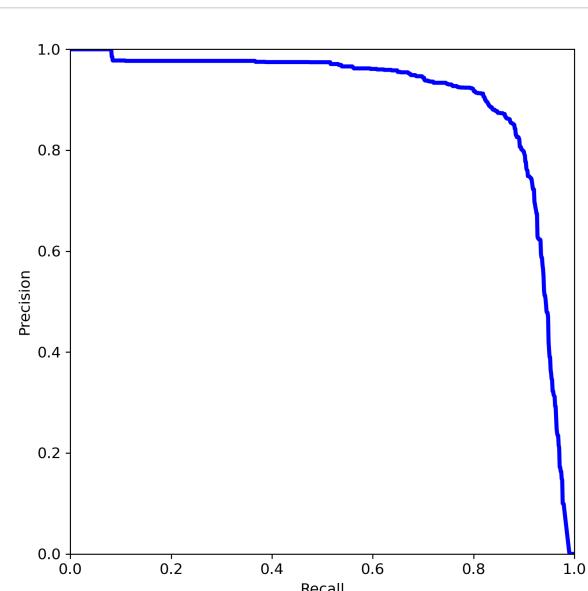
FIGURE 8
Schematic diagram of PR curve.
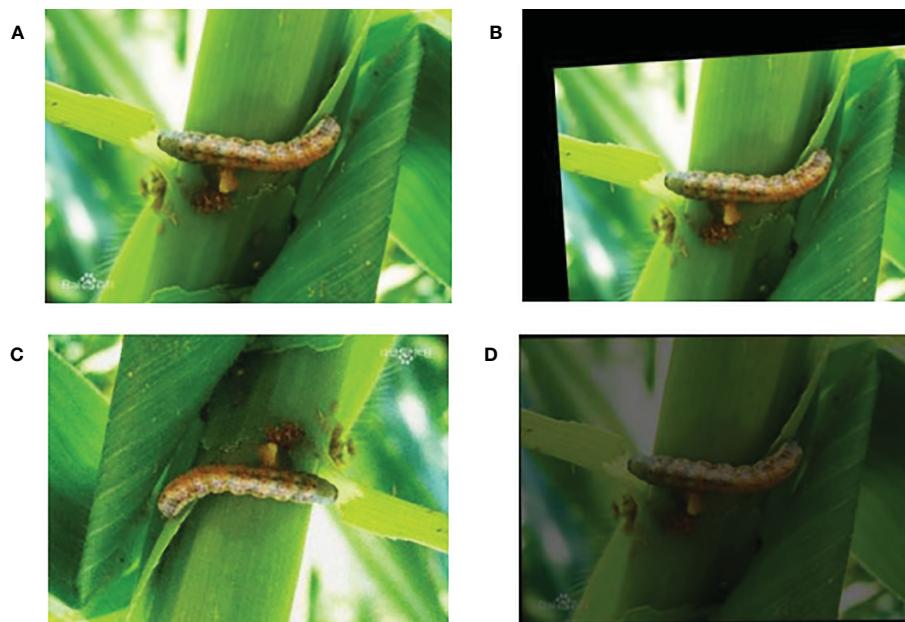
FIGURE 9
Data augmentation effect display: **(A)** Original image. **(B)** Random crop. **(C)** Flip. **(D)** Decrease in brightness.

3.4 Data augmentation

As deep learning requires a large amount of data for training, we used data augmentation on the original dataset, such as random rotation transform, blur transformation, flip transform, addition of Gaussian noise and so on. The random rotation and flip transformation models are able to simulate the different locations of insect presence, while the blur transformation and Gaussian noise could better simulate the various environment that may occur in reality. Figure 9 shows the images which performing data augmentation.

3.5 Transfer learning

Transfer learning is a popular method in the field of computer vision, because it can build accurate models in less time. By using transfer learning, model do not start training from scratch, but start with the patterns of solving problems that learned from previous problems. In the field of computer vision, transfer learning is usually represented by the use of pre-trained models. Pre-trained models are models that trained on large baseline datasets. For example, in object detection tasks, backbone neural network is first used for feature extraction. The backbone used here is generally a neural network such as VGG, ResNet, etc. Therefore, when training an object detection model, the parameters of the backbone can be initialized by using the pre-trained weights of these neural networks so that more effective features can be extracted at the beginning.

In this paper, we selected the IP102 public dataset as a pre-trained dataset[1]. The IP102 dataset is a large-scale dataset for pests identification, which contains more than 75,000 images of 102 pests classes. These images exhibit a natural long-tailed distribution. In addition, about 19,000 of these images have added bounding boxes for object detection. We select these images with object detection frames, and feed them into individual networks for training to obtain pre-trained weights. The pre-trained weights will be transferred to our own dataset for use, and it can make the final model more robust and convincing in the pests identification task.

3.6 Experimental environment and parameter settings

The experimental environment configuration of this paper is as follows: OS is Linux, GPUs are two Tesla V100 with 80G memories, training environment is python 3.7, Pytorch 1.11.0. while Labelme is used to annotate the data. In training, to ensure comparability across experiments and appropriateness of training, each training epoch consists of 100 rounds and the img_size is 320×320. In order to verify the good performance of our proposed algorithm under large batch size, we set the batch size to 512. While for training of YOLOv7, the weight_decay is 0.002 and learning rate is 0.001.

1   https://github.com/xpwu95/IP102

3.7 Experiment result

Figure 10 shows the prediction performance of the YOLOv7 network incorporating with the Adan optimizer when facing different species of maize pests.

In order to verify the effectiveness of the algorithm proposed in this paper, we compared the improved network with the original network which using Adam, AdamW and SGD. We also tested several other object detection networks: SSD (Zhao et al., 2015), RetinaNet (Zhu et al., 2010), FCOS (Zhu et al., 2012), Faster RCNN (Xu et al., 2002) and FPN (Zhu and Zhang, 2013). Finally, we put these networks together and compared them with the result of our previous works, and the performance evaluation indexes are [mAP@.5:.95] and precision which are described above. The result is shown in Table 2.

We also compared the differences between the YOLOv7 network loaded with Adan and other networks when face with the same image. And the prediction result are shown in Figure 11.

4 Discussion

The experimental result in Table 2 shows that the YOLOv7 network incorporating the Adan outperforms traditional ML algorithms and other comparative networks in the comparison of both mAP@[.5:.95] and precision. Meanwhile, from Figure 10 we can see that the improved network has a good performance on different types of maize pests. What's more, further comparison of three different networks in Figure 11 shows that the YOLOv7 network incorporating Adan can still perform well in more complex natural environment with no errors. SSD network and the YOLOv7 network incorporating the Adam both have errors in prediction of the same images. The YOLOv7 network with the Adam misidentified the background as insects in two images, while SSD network misidentified insects as the background in both images. The final comparison of performance indexes and prediction result verifies that Adan optimizer can effectively improve the model performance and help the YOLOv7 network


FIGURE 10
Test result: **(A)** bollworm. **(B)** armyworm. **(C)** corn borer.

TABLE 2 Performance comparison of different networks.

Networks	mAP@[.5:.95]	Precision
YOLOv7(Adam) (Vızhányó and Felföldi, 2000)	0.8646	99.66%
YOLOv7(AdamW) (Vızhányó and Felföldi, 2000)	0.9032	99.64%
YOLOv7(SGD) (Vızhányó and Felföldi, 2000)	0.9407	99.91%
Faster R-CNN (Xu et al., 2002)	0.6655	88.99%
SSD (Zhao et al., 2015)	0.6608	98.87%
RetinaNet (Zhu et al., 2010)	0.681	98.2%
FCOS (Zhu et al., 2012)	0.602	86.8%
FPN (Zhu and Zhang, 2013)	0.6337	87.6%
Histogram Reverse Mapping and Invariant Moment (Redmon and Farhadi, 2018)	None	94.3%
GMM and DRLSE (Ren et al., 2015)	None	86.364%
Ours	**0.9669**	**99.95%**

The bold values means the better results.

FIGURE 11
Comparison of the prediction effect of different networks when facing the same image: **(A)** YOLOv7(Adan). **(B)** YOLOv7(Adam). **(C)** SSD.

reduce the possibility of false recognition and missed recognition, thus making the network more efficient and error-free in pests recognition task. For further confirmation, we collected data of the map[.5:.95] and precision of YOLOv7 networks which using different optimizers in the experiment when the epoch changed, as shown in Tables 3, 4. Based on these data, we plotted the performance trends of four optimizers, as shown in Figure 12.

From Figure 12 we can see that the YOLOv7 incorporating with Adan optimizer converges faster than YOLOv7 loaded with other optimizers in both mAP@[.5:.95] and precision, and the result are consistent with our theoretical analysis. In process of calculating momentums, Adan uses the modified Nesterov momentum algorithm, while Adam with AdamW use the traditional reballing algorithm. The modified Nesterov momentum algorithm helps Adan to sense the surrounding gradient information in advance, which helps model to escape from the sharp local minimal regions efficiently, thus speeding up the convergence of Adan. The comparison of map[.5:.95] and precision shows that Adan can obtain greater performance by using only 1/2-2/3 of the computation of other optimizers. What's more, the mAP@[.5:.95] increases by 2.79%-11.83% compared to original optimizers. The experimental result also confirm that Adan only needs less than 2/3

of computation of the original network to obtain the performance beyond it, which is proposed in the original paper of Adan.

5 Conclusions

In this paper, a new deep learning algorithm based on YOLOv7 network and Adan optimizer is proposed, and a feasible maize pests identification scheme is proposed as well, which is successfully applied to the identification task of maize pests. The mAP@[.5:.95] of the improved network reaches 96.69% and the precision reaches 99.95% in this task, which breaks the bottleneck of the original networks. And it also confirms the feasibility and effectiveness of applying deep convolutional neural networks to the task of crop pests and disease identification, and it has positive significance for crop pests and disease prevention and control. We can quickly identify common corn pests and take appropriate measures by using this model, and scientifically carryout pests control methods to reduce possible economic losses and promote agricultural modernization.

However, the environment is more complex in real life. There are many other insects with similar characteristics, while the difficulty of detection in complex environment will be greatly increased due to the

TABLE 3 Comparison of [mAP@.5:.95] changes with epochs for different optimizers.

epochs	Ours	YOLOv7(Adam)	YOLOv7(AdamW)	YOLOv7(SGD)
10/100	**0.4202**	0.342	0.2719	0.241
20/100	**0.5444**	0.2034	0.3086	0.3765
30/100	**0.7107**	0.4091	0.4758	0.3746
40/100	**0.8681**	0.5632	0.7297	0.5812
50/100	**0.8691**	0.7674	0.802	0.6337
60/100	**0.9212**	0.8176	0.8966	0.7758
70/100	**0.9401**	0.8409	0.8846	0.8022
80/100	**0.9615**	0.8699	0.9083	0.8103
90/100	**0.9658**	0.9089	0.9289	0.8916
100/100	**0.9669**	0.8646	0.9032	0.9407

The bold values means the better results.

TABLE 4 Comparison of precision changes with epochs for different optimizers.

epochs	Ours	YOLOv7(Adam)	YOLOv7(AdamW)	YOLOv7(SGD)
10/100	0.3589	**0.4026**	0.2933	0.4015
20/100	**0.8744**	0.3229	0.5222	0.5031
30/100	**0.968**	0.6481	0.7153	0.6272
40/100	0.9717	0.7398	0.9332	**0.9803**
50/100	**0.9974**	0.9466	0.9774	0.995
60/100	**0.9981**	0.9823	0.9962	0.9979
70/100	**0.9986**	0.9825	0.9978	0.9967
80/100	**0.9991**	0.9918	0.9972	0.9989

(Continued)

TABLE 4 Continued

epochs	Ours	YOLOv7(Adam)	YOLOv7(AdamW)	YOLOv7(SGD)
90/100	0.999	0.9959	0.997	**0.9997**
100/100	**0.9995**	0.9966	0.9982	0.9991

The bold values means the better results.
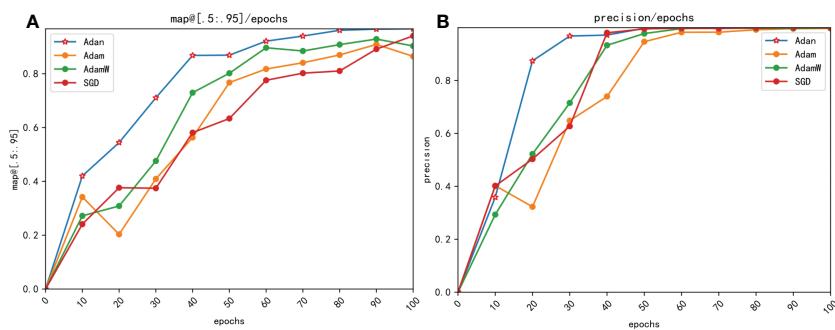
FIGURE 12
Comparison of precision changes with epoches for different optimizers. **(A)** map@[.5:.95] **(B)** precision.

limitations of scarce data. Meanwhile, some corn pests will appear in the form of eggs in real life, while these eggs are tiny and their characteristics are difficult to distinguish, making identification more difficult. What's worse, pests data are scarce and difficult to collect, and the cost of manual labeling is very high. Therefore, how to obtain sufficient data and enough computing power is the key of future pests controlling technology researches.

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

Conceptualization, ZH, YZ and CZ; methodology, CZ and LW; software, CZ and LW; validation, ZH, CZ, and LW; formal analysis, ZH, LW, and CZ; investigation, ZH and YZ; resources, ZH; data curation, CZ and LW; writing—original draft preparation, CZ and LW; writing—review and editing, ZH; visualization, CZ; supervision, ZH; project administration, ZH and YZ; funding acquisition, ZH and YZ. All authors contributed to the article and approved the submitted version.

Funding

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

Al Bashish, D., Braik, M., and Bani-Ahmad, S. (2010a). "A framework for detection and classification of plant leaf and stem diseases," in *2010 international conference on signal and image processing (IEEE)*. 113–118.

Amara, J., Bouaziz, B., and Algergawy, A. (2017). "A deep learning-based approach for banana leaf diseases classification," in Datenbanksysteme für Business, Technologie und Web (BTW 2017), 17. Fachtagung des GI-Fachbereichs, Datenbanken und Informationssysteme (DBIS). Stuttgart, Germany: Workshopband, 6–10.

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). *Yolov4: optimal speed and accuracy of object detection* (arXiv preprint arXiv:2004.10934).

Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*. 2961–2969.

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., et al. (2022). *Yolov6: a single-stage object detection framework for industrial applications* (arXiv preprint arXiv:2209.02976).

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*. 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*. 2980–2988.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: single shot multibox detector," in *European Conference on Computer Vision* (Amsterdam, The Netherlands: Springer), 21–37.

Liu, B., Hu, Z., Zhao, Y., Bai, Y., and Wang, Y. (2019). *Recognition of pyralidae insects using intelligent monitoring autonomous robot vehicle in natural farm scene* (arXiv preprint arXiv:1903.10827).

Liu, F., Shen, Z., Zhang, J., and Yang, H. (2008). Automatic insect identification based on color characters. *Chin. Bull. Entomol.* 45, 150–153. doi: 10.1016/S1005-9040(08)60003-3

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 1419. doi: 10.3389/fpls.2016.01419

Nachtigall, L. G., Araujo, R. M., and Nachtigall, G. R. (2016). "Classification of apple tree disorders using convolutional neural networks," in *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI) (IEEE)*. 472–476.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

Redmon, J., and Farhadi, A. (2017). "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7263–7271.

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

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

Sammany, M., and Medhat, T. (2007). "Dimensionality reduction using rough set approach for two neural networks-based applications," in *Rough Sets and Intelligent Systems Paradigms*. (Warsaw, Poland: Springer), 639–647.

Sasaki, Y., Okamoto, T., Imou, K., and Torii, T. (1998). Automatic diagnosis of plant disease-spectral reflectance of healthy and diseased leaves. *IFAC Proc. Volumes* 31, 145–150. doi: 10.1016/S1474-6670(17)42113-6

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* 2016, 1–11. doi: 10.1155/2016/3289801

Tian, Z., Shen, C., Chen, H., and He, T. (2019). "Fcos: fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*. 9627–9636.

Vızhányó, T., and Felföldi, J. (2000). Enhancing colour differences in images of diseased mushrooms. *Comput. Electron. Agric.* 26, 187–198. doi: 10.1016/S0168-1699(00)00071-5

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). *Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors* (arXiv preprint arXiv:2207.02696).

Wang, J., Zhao, Y., Wang, Y., Chen, W., Li, H., Han, Y., et al. (2020). "Marked watershed algorithm combined with morphological preprocessing based segmentation of adherent spores," in *International Conferences on Communications, Signal Processing, and Systems*. (Springer), 1316–1323.

Xie, X., Zhou, P., Li, H., Lin, Z., and Yan, S. (2022). *Adan: adaptive nesterov momentum algorithm for faster optimizing deep models* (arXiv preprint arXiv:2208.06677).

Xu, L., Mao, H., and Pingping, L. (2002). Study on color feature extraction of color images of vegetation-deficient leaves. *J. Agric. Eng.* 4, 150–154. doi: 10.3321/j.issn:1002-6819.2002.04.037

Zhao, Y., and Hu, Z. (2015). Segmentation of fruit with diseases in natural scenes based on logarithmic similarity constraint otsu. *Trans. Chin. Soc. Agric. Machinery* 46, 9–15. doi: 10.6041/j.issn.1000-1298.2015.11.002

Zhao, Y., Hu, Z., Bai, Y., and Cao, F. (2015). An accurate segmentation approach for disease and pest based on drlse guided by texture difference. *Trans. Chin. Soc. Agric. Machinery* 46, 14–19. doi: 10.6041/j.issn.1000-1298.2015.02.003

Zhao, Y., Liu, S., Hu, Z., Bai, Y., Shen, C., and Shi, X. (2020). Separate degree based otsu and signed similarity driven level set for segmenting and counting anthrax spores. *Comput. Electron. Agric.* 169, 105230. doi: 10.1016/j.compag.2020.105230

Zhao, Y., Wang, Y., Wang, J., Hu, Z., Lin, F., and Xu, M. (2019). "Gmm and drlse based detection and segmentation of pests: a case study," in *Proceedings of the 2019 4th International Conference on Multimedia Systems and Signal Processing*. 62–66.

Zhao, Y., Wang, K., Zhongying, B., Li, S., Xie, R., and Gao, S. (2007). Application of bayesian methods in image recognition of maize leaf diseases. *Comput. Eng. Appl.* 5, 193–195. doi: 10.3321/j.issn:1002-8331.2007.05.058

Zhu, L.-Q., and Zhang, Z. (2013). Using cart and llc for image recognition of lepidoptera. *Pan-Pacific Entomol.* 89, 176–186. doi: 10.3956/2013-08.1

Zhu, L., Zhang, Z., et al. (2012). Automatic insect classification based on local mean colour feature and supported vector machines. *Oriental Insects* 46, 260–269. doi: 10.1080/00305316.2012.738142

Zhu, L., Zhang, Z., Zhang, P., et al. (2010). Image identification of insects based on color histogram and dual tree complex wavelet transform (dtcwt). *Acta Entomol. Sin.* 53, 91–97. doi: 10.16380/j.kcxb.2010.01.016

Zhu, L., Zhang, D., Zhang, Z., et al. (2015a). Feature description of lepidopteran insect wing images based on wld and hoc and its application in species recognition. *Acta Entomol. Sin.* 58, 419–426.

Zhu, L., Zhang, D., Zhang, Z., et al. (2015b). Recognition of lepidopteran species based on color name and opponentsift features. *Acta Entomol. Sin.* 58, 1331–1337. doi: 10.16380/j.kcxb.2015.04.008

Check for updates

Research of segmentation recognition of small disease spots on apple leaves based on hybrid loss function and CBAM

Xiaoqian Zhang†, Dongming Li†, Xuan Liu, Tao Sun,
Xiujun Lin and Zhenhui Ren*

College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, China

Identification technology of apple diseases is of great significance in improving production efficiency and quality. This paper has used apple Alternaria blotch and brown spot disease leaves as the research object and proposes a disease spot segmentation and disease identification method based on DFL-UNet+CBAM to address the problems of low recognition accuracy and poor performance of small spot segmentation in apple leaf disease recognition. The goal of this paper is to accurately prevent and control apple diseases, avoid fruit quality degradation and yield reduction, and reduce the resulting economic losses. DFL-UNet +CBAM model has employed a hybrid loss function of Dice Loss and Focal Loss as the loss function and added CBAM attention mechanism to both effective feature layers extracted by the backbone network and the results of the first upsampling, enhancing the model to rescale the inter-feature weighting relationships, enhance the channel features of leaf disease spots and suppressing the channel features of healthy parts of the leaf, and improving the network's ability to extract disease features while also increasing model robustness. In general, after training, the average loss rate of the improved model decreases from 0.063 to 0.008 under the premise of ensuring the accuracy of image segmentation. The smaller the loss value is, the better the model is. In the lesion segmentation and disease identification test, MIoU was 91.07%, MPA was 95.58%, F1 Score was 95.16%, MIoU index increased by 1.96%, predicted disease area and actual disease area overlap increased, MPA increased by 1.06%, predicted category correctness increased, F1 Score increased by 1.14%, the number of correctly identified lesion pixels increased, and the segmentation result was more accurate. Specifically, compared to the original U-Net model, the segmentation of Alternaria blotch disease, the MIoU value increased by 4.41%, the MPA value increased by 4.13%, the Precision increased by 1.49%, the Recall increased by 4.13%, and the F1 Score increased by 2.81%; in the segmentation of brown spots, MIoU values increased by 1.18%, MPA values by 0.6%, Precision by 0.78%, Recall by 0.6%, and F1 Score by 0.69%. The spot diameter of the Alternaria blotch disease is 0.2-0.3cm in the early stage, 0.5-0.6cm in the middle and late stages, and the spot diameter of the brown spot disease is 0.3-3cm. Obviously, brown spot spots are larger than Alternaria blotch spots. The segmentation performance of smaller disease spots has increased more noticeably, according to the quantitative analysis results, proving that the model's capacity to segment smaller disease spots has greatly improved. The

findings demonstrate that for the detection of apple leaf diseases, the method suggested in this research has a greater recognition accuracy and better segmentation performance. The model in this paper can obtain more sophisticated semantic information in comparison to the traditional U-Net, further enhance the recognition accuracy and segmentation performance of apple leaf spots, and address the issues of low accuracy and low efficiency of conventional disease recognition methods as well as the challenging convergence of conventional deep convolutional networks.

Introduction

Apples are rich in medicinal and nutritional value and are one of the most widely planted fruit industries in the world (Khan et al., 2022). From the data of recent years, the growth rate of apple production has been decreasing year by year (Liu et al., 2018), and analyzing the reasons for this, diseases are one of the important influencing factors. Diseases of apple trees occur in the roots, branches, fruits, and leaves, and most of them initially spread from the leaves, so accurate and effective identification of apple leaf disease types and the degree of disease is a key aspect of apple disease protection management. According to statistics, there are more than 100 kinds of apple leaf diseases, among which Alternaria blotch and brown spot disease are the two most common leaf diseases of apple trees. In this paper, we have segmented the spots and classified the diseases for the 2 common types of apple leaf diseases.

The traditional method of judging fruit tree leaf diseases mainly relies on expert experience by manually extracting the color, texture, and shape characteristics of diseased leaf images (Ayaz et al., 2021). However, in actual production activities, it is easy to misjudge the type of disease and thus misuse pesticides, which affects apple production. Therefore, a more convenient and accurate disease diagnosis method is urgently needed to analyze and determine the type of disease which provides researchers with a reasonable application strategy to prevent and control the disease on time and reduce the planting management pressure of fruit farmers. With the breakthrough progress of deep convolutional neural networks in classification tasks on open data sets, many scholars have applied image segmentation technology to the field of disease spot recognition to segment disease images and identify them in real-time, scientifically determine the type of leaf diseases and the degree of disease, take timely and effective measures to improve apple yield, and help fruit farmers achieve early disease control.

The current challenges of apple leaf and spot image segmentation can be summarized into the following three types:

1. Unbalanced pixel ratio. The disease spot information is readily lost in the disease spot segmentation task because the pixels in the diseased region only make up a small portion of all the pixels in the entire image. Additionally, because of the imbalanced pixel ratio, a lot of pixels in the background that can be classified easily hide a lot of the pixels in the rare diseased zone during the loss summing, which negatively affects model training and, as a result, the segmentation of diseased spots.

2. Hard example sample problem. The extraction of target leaf edges and disease areas is problematic in the natural environment due to leaf overlap, uneven lighting, and shadows. These difficult-to-classify pixels directly affect the outcomes of leaf segmentation, which in turn affects the extraction of disease spots.

3. When an apple tree is infected in its early stages, the fruit has not yet developed, and the illness first appears in the leaves. Brown to dark brown little round spots with a diameter of 2 to 3 mm was generated on the young leaves during the early stages of spotted defoliation, and purple haloes were frequently present surrounding the lesions with obvious margins. Yellowish-brown dots that eventually became circular emerged on the leaf surface in the early stages of brown spot disease. The early stages of spotted defoliation and brown spot are quite similar, making it challenging to tell them apart. This makes it difficult to identify the types of diseases, which has an impact on early disease prevention and control.

In order to more precisely locate disease areas and identify disease species, as well as to lay the groundwork for future assessments of the severity of disease in fruit trees and effective disease control methods, the main motivation for the current study is to segment the smaller spots on apple leaves and classify similar diseases. Smaller spots are challenging to identify in lesion segmentation, necessitating model improvement to enhance lesion segmentation performance. Early detection of apple leaf diseases is essential for timely disease management, illness prevention, and mitigation of effects on fruit quality and fruit yield. Further, the performance of various semantic segmentation models (such as Deeplabv3+, PSPNet (Pyramid Scene Parseing Network), and U-Net) in spot segmentation has been the focus of recent research, and performance evaluation measures like MIOU

(Mean Intersection over Union), MPA (Mean Pixel Accuracy), Precision, and F1 scores were taken into account in the current work.

Too et al. (Too et al., 2019) compared various convolutional neural network models, including VGG16, InceptionV4, ResNet50, ResNet101, ResNet152, and DenseNets121, using plant leaf diseases from the publicly available Plant Village dataset as the research object. The results of the experiment revealed that the DenseNets network model performed the best in terms of classifying and identifying plant leaf diseases. Lin et al. (Lin et al., 2019) improved the UNet-based deep convolutional neural network model was proposed for cucumber powdery mildew to segment and extract the diseased areas of cucumber leaves with an average pixel accuracy of 96.08%, which is better than traditional detection methods such as K-means, random forest, and GBDT (Gradient Boosting Decision Tree). Zhong Y et al. (Zhong and Zhao, 2020) proposed three methods to identify apple leaf diseases: regression, multi-label classification, and Focal Loss function based on DenseNet-121 deep convolutional network, and the accuracy of the method on the test set was 93.51%, 93.31%, and 93.71%, respectively, which was better than the traditional CE (cross-entropy) loss function-based multi-classification methods. Santos et al. (Santos et al., 2020) used the Mask R-CNN instance segmentation network model to segment, detect and count the grape trees in the real scene, compared with other network models, the F-score of the Mask R-CNN network model achieved an optimal effect of 0.91. Ngugi et al. (Ngugi et al., 2020) modified the encoder component of the UNet network model to offer the network the ability of multi-scale feature extraction to achieve tomato leaf disease spot segmentation on complicated backdrops, thus increasing the segmentation accuracy of tomato leaf illnesses. On the entire plant leaf specimen dataset, Hussein et al. (Hussein et al., 2021) used DeepLabV3+ to conduct segmentation experiments and found that utilizing a deep learning semantic segmentation model produced superior semantic segmentation outcomes than target detection techniques like Faster R-CNN (Ren et al., 2017) and Yolo v5. Wang P et al. (Wang et al., 2021) proposed to use CA-ENet to identify different apple diseases. This method integrates a coordinate attention block in the EfficientNet-B4 network, uses deep separable convolution in the convolution module, and introduces the h-swish activation function. The experimental results show that the accuracy of this method is 98.92%, and the average F1 score is 0.988, which is better than ResNet-152, DenseNet-264, and ResNeXt-101. Tassis L M et al. (Tassis et al., 2021) used the Mask R-CNN network, U-Net, and PSPNet networks to automatically detect identify disease spots in field images containing some coffee trees and obtained 73.90% accuracy and 71.90% recall in the instance segmentation task; for U-Net and PSPNet networks, 94.25% and 93.54% average intersection and union were obtained. Li X et al. (Li et al., 2022) used U-Net, PSPNet, and DeepLabV3+ (Chen et al., 2018a) semantic segmentation model for potato leaf segmentation, and the MIoU of the model was 89.91% and MPA was 94.24%.

Studies have shown that plant leaf lesion segmentation based on deep learning semantic segmentation models is feasible, but existing studies have only used CNN-based models to identify crops and plant diseases without improving the models, and there are fewer studies on segmentation of apple leaf lesion regions based on semantic segmentation. Liu et al. (Liu et al., 2022) used the severity of apple Alternaria blotch assessed using DeeplabV3 +, PSPNet, and UNet. The correlation coefficient and consistency correlation coefficient were both 0.992 and the average accuracy of severity categorization was 96.41%. The study's lack of many disease instances in a single leaf image was a drawback, even though the reference value and anticipated value were in agreement. In addition, in prior research, the loss function of the model is typically a single loss function. In this study, to enhance the segmentation performance and achieve more precise segmentation of leaves and disease spots under natural conditions, we fused two loss functions and added attention mechanisms to both the two effective feature layers extracted by the backbone network and the outcomes of the first upsampling.

Therefore, this paper has improved the U-Net model by adopting a hybrid loss function and adding an attention mechanism to perform pixel feature extraction and spot segmentation for two common types of apple leaf diseases, so that the disease can be recognized accurately. This method has improved the recognition accuracy and segmentation effect for small targets such as apple leaf spots while ensuring its feature extraction and classification ability.

The main contributions of this work are as follows:

1. Dice Loss and Focal Loss are combined as the loss function in this paper, causing the network to pay more attention to the similarity of lesions, increase the accuracy of image segmentation, and optimize the segmentation details.
2. The original U-Net model is proposed to be enhanced with an attention mechanism in this research. By comparing the segmentation accuracy after incorporating the three attention mechanisms SENet (Squeeze-and-Excitation Networks), ECANet (Efficient Channel Attention Module), and CBAM (Convolutional Block Attention Module), it is found that adding CBAM to the original model improves the network's capacity to extract illness features and increases the robustness of the model.
3. The model in this work has the best segmentation performance in smaller disease spots segmentation recognition when the segmentation performances of Deeplabv3+, PSPNet, U-Net, and DFL-UNet+CBAM are compared.
4. The classification and identification of related diseases, as well as the segmentation and recognition of smaller disease spots in apple leaves, were accomplished. In general, the results of this experiment can serve as a technical foundation for the future segmentation, classification, and precise management of plant leaf disease spots.

The structure of the whole document is as follows. The first section of this essay provides an overview of the study context and topic's importance, the research's driving forces, its current state, its main contributions, and its primary ideas. In Section 2, the suggested modeling strategy is introduced, with an emphasis on

the U-Net algorithm, the attention mechanism, and the loss function, as well as a description of the enhanced network topology. The study's materials and procedures are described in Section 3, including the dataset preparation process, model training environment, model implementation platform, and an explanation of each model assessment metric's parameter. The fourth section examines the experimental findings, investigates the segmentation impact of the model trained on smaller disease spots using a variety of algorithms, loss functions, and attention processes, and discusses the training strategy for the model with the best segmentation effect. The discussion of the research is introduced in Section 5, which mostly outlines the issues that need to be resolved. Section 6 summarizes the research of this paper and introduces the research conclusions of the test.

Improved U-Net network structure

U-Net network structure

One of the earliest full convolutional network-based image segmentation algorithms, U-Net is an upgraded semantic segmentation network built on FCN (Fully Convolutional Networks) (Shelhamer et al., 2017) and may maintain more local features in the segmentation outcomes.

The "U-shaped" symmetric encoder-decoder structure of the U-Net network's second half, which is upsampling, is used for feature extraction in the first half. The enhanced feature extraction part of the process can be used to up-sample the five initial effective feature layers obtained from the backbone part and perform feature fusion to obtain an effective feature layer that fuses all features to classify each feature point. The backbone feature extraction part makes up the first half.

Loss function

In this paper, a hybrid loss function was utilized to close the gap between the prediction results and the true values and achieved high confidence in the boundaries of segmented images. The commonly used loss function was CE Loss, but its role was relatively small when the examples were unbalanced. Lin, T.-Y. et al. (Lin et al., 2020) proposed focal loss to improve the accuracy of dense object detection. Dice Loss (Wang et al., 2020) and Focal Loss (Chen and Qin, 2022) were taken into consideration in order to address the issues of poor segmentation performance of smaller disease spots in apple leaves and the challenge of classifying apple Alternaria blotch and brown spot disease diseases with similar disease characteristics at the early stage of disease onset.

The basic idea behind Dice Loss was to measure the regional similarity between the prediction result and the true value during training; however, using Dice Loss directly reduced training stability. To avoid the problem of assigning different weights to the same class while ignoring the presence of hard examples in both positive and negative examples, such as pixels in the diseased area covered by raindrops and light or other leaf pixels in the background, the network was focused on learning hard examples

by using a Focal Loss function that increases the loss value of challenging examples. By increasing the loss value of hard examples and forcing the network to concentrate on learning hard examples, it addressed the issue of unbalanced positive and negative examples as well as unbalanced hard and easy examples.

(1) CE loss is used to measure the difference between two probability distributions and the gap between model learning and reality. The traditional cross-entropy loss function is the most often used loss function in classification. Equation (1) displays its formula.

$$CE \ Loss = -(y_i \log p_i + (1 - y_i) \log (1 - p_i)) \tag{1}$$

(2) Dice Loss places more emphasis on identifying leaf regions and gauges how well the outcomes anticipated and actual values in the area match up. Equation (2) illustrates the formula.

$$Dice \ loss = \frac{2TP}{2TP + FN + FP} \tag{2}$$

where, correspondingly, TP (True Positive), FP (False Positive), and FN (False Negative) represent the total number of true positives, false positives, and false negatives.

(3) Focal Loss focuses the network on learning hard examples by enhancing the loss value of hard examples, balancing positive and negative examples and difficult and easy classification examples, as shown in equation (3).

$$Focal \ loss(Y, P) = -\frac{1}{n} \sum_{i=1}^{n} [\alpha y_i (1 - p_i)^\gamma \ln p_i + (1 - \alpha)(1 - y_i) p_i^\gamma \ln (1 - p_i)] \tag{3}$$

In the equation, n stands for the total number of apple leaf samples, y_i for the input sample's true category, p_i for the likelihood that the sample is 1, and γ for the modulation coefficient. The average logarithmic loss for each sample is shown by the logarithmic loss for all samples. To strengthen the focus on positive examples and improve the imbalance of targets in the case of extremely unbalanced categories, adding $(1 - p_i)^\gamma$ will cause the loss value of samples with high prediction probability to decrease while the loss value of samples with low prediction probability to increase. Currently, image segmentation can only use it for binary classification. The positive example in the binary classification problem has a label of 1, and the negative example has a label of 0. For the positive example, the more $1 - p_i$, the harder it is to categorize the sample. The more p_i is greater, the more challenging it is to classify negative examples.

In this study, the loss function employed a hybrid loss function (DFL) that scaled both Dice Loss and Focal Loss to the same order of magnitude to predict the input data, with Dice Loss emphasizing similarity and Focal Loss improving segmentation specifics to increase image segmentation accuracy.

Attentional mechanism

Jain et al. (Jain et al., 2022) compared the Attention-UNet model with the UNet, UNet + + and UNet3P models, the AUC (Area Under Curve) value is 0.97, while the AUC values of other

models are 0.964,0.966 and 0.965, respectively. The results show that the attention mechanism is beneficial to segment very bright and blurred plaque images that are difficult to diagnose using other methods.

The inclusion of an attention mechanism was thought to improve feature extraction because the leaf spot areas are smaller. The model would then assign different weights to each location of the input image and concentrate on the areas with crucial information, which would help it make more accurate judgments while using fewer resources. The attention mechanism has demonstrated strong performance in previous research on tasks like categorization, detection, and segmentation (Karthik et al., 2020; Mi et al., 2020). In this study, we thoroughly examined SENet (Hu et al., 2020), ECANet (Yu et al., 2022), and CBAM (Ma et al., 2022), three attention mechanisms, and we chose the best module to enhance apple leaf spot segmentation.

ECANet removed the two FC (Fully Connected) layers used in SENet and performed global average pooling without dimensionality reduction. It used the current channel and its k neighboring channels for local cross-channel interaction. SENet performed global average pooling of the input feature layer, took the Sigmoid after completing two full joins, obtained the weight of each channel of the input feature layer, and then multiplied that weight by the original input feature layer. Compared to SENet's attention mechanism, which focused exclusively on channels, CBAM was a lightweight attention module that could be integrated into virtually any convolutional neural network, and almost negligible computation and parameters were introduced. It combined the channel attention mechanism and the spatial attention mechanism to jointly learn the important local detail information in the image, assign higher weights to the diseased spot region in the neural network's feature map and lower weights to the background, improved the neural network's attention to the diseased spot in the image, and then enhanced the network's capacity for feature learning and expression.

In order to boost the network's capacity to extract disease features and the resilience of the model, an attention mechanism was added to the two effective feature layers that the backbone network extracted, as well as to the outcomes of the initial upsampling.

Improved U-Net network structure

This paper proposed an improved model based on U-Net that keeps the backbone feature extraction network but enhanced it by adding CBAM modules to the two effective feature layers extracted by the backbone network; after being subjected to feature fusion to complete two convolution operations, the effective feature layers obtained in the coding stage are then subjected to upsampling to recover the original image accuracy and detail information pixel by pixel. The CBAM attention module was then embedded after the first upsampling. The model was designed to recalibrate the weight relationships between features, amplify channel features of leaf disease spots, and suppressed channel features of healthy regions of leaves to improve the network's ability to extract disease features

and to increase the model's robustness. The upper portion of the improved U-Net network was the backbone feature extraction network, and the lower portion was the enhanced feature extraction network.

Additionally, the improved model predicts the input data using a mixed loss function (DFL), which scales the focus loss and dice loss to the same magnitude. During training, Dice Loss focuses more on identifying the foreground region and assesses how closely the results of the prediction match the actual value in the area. By strengthening the loss value of hard examples (such as pixels in the diseased area covered by raindrops and light or other leaf pixels in the background), the Focal Loss function makes the network focus on the learning of difficult samples and solves the problem of imbalance between positive and negative examples and imbalance between difficult and easy samples. The structure of the network is shown in Figure 1.

Materials and methods

Dataset source

The image samples of apple leaf diseases in this experiment came from the public data set Plant Village (Geetharamani and Pandian, 2019). The dataset manually collects images of indoor and outdoor diseased apple leaves. In order to ensure the versatility of the model, outdoor landscape images were taken on sunny and rainy days, respectively.

The data set in this paper contains different situations of a single leaf with a single disease and multiple diseases and multi-leaf images in complex backgrounds. As seen in Figure 2, the leaf diseases include single Alternaria blotch, brown spot, and multiple diseases (including brown spot and mosaic) of apple leaves. The samples in this dataset include pre-processing operations on the acquired images, such as image rotation, horizontal and vertical mirroring, a sharpness value, brightness value, contrast adjustment, and Gaussian blur on the original disease images. This pre-processing was done to prevent overfitting issues in the later network training phase, to improve the anti-interference ability of complex conditions as well as the generalization ability of the model, to increase the diversity, and to avoid generating problems during the network training phase, and thus the model robustness is enhanced.

Also to ensure a balanced sample, 1200 images of a single Alternaria blotch, 1200 images of a single brown spot, 600 images of apple leaf diseases infected with multiple diseases (mosaic and brown spot), a total of 3000 original images (JPG format) were selected, with a 1:1 ratio of complex background images to simple laboratory background images, which is more challenging than laboratory images of diseased leaves with simple backgrounds, with an original image size of 512 pixels * 512 pixels, and the dataset was divided into training, validation and test sets in a 6:2:2 ratio. As demonstrated in Figure 2, the image of apple leaf disease has the traits of a smaller disease spot and high similarity, which presents numerous difficulties for image segmentation.

For leaf segmentation, it is difficult to extract the target leaf's edge because there are multiple leaves overlapping in the
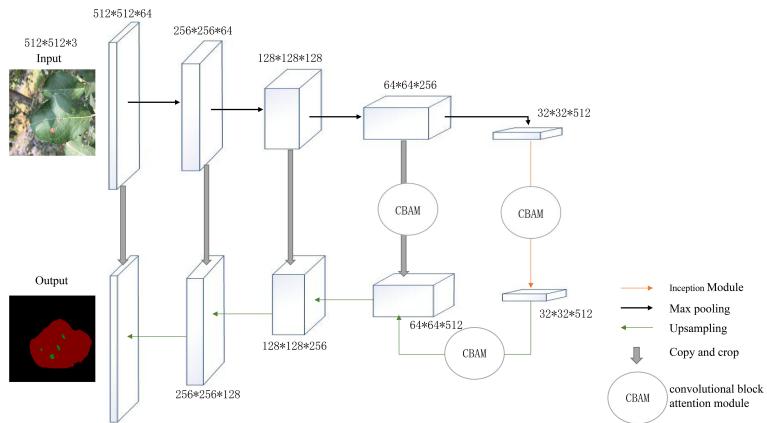
FIGURE 1
Improved U-Net network structure diagram.

background of the image in the outdoor scene. Additionally, there are shadows in the leaf image due to uneven illumination, self-crimping, folding, and other factors, which makes the segmentation more challenging. Diverse and challenging to extract features from the target leaf scales. The extraction of disease spot features and the precise detection of disease spots are greatly hampered by the smaller disease spot pixels, which make up 0.2% to 0.4% of the leaf pixels in spot segmentation. Outside, there are materials that resemble spots that could prevent infections from being extracted. The segmentation impact of disease spots is easily influenced by the spots on raindrops and leaves.

Dataset production

The photos must be converted into a dataset in PASCAL VOC format by the specifications of the model for the dataset. JPEGImages, ImageSets, and Annotations were the three main files that made up the PASCAL VOC format dataset.

The Segmentation folder of the ImageSets file contained four text files: train.txt, val.txt, test.txt, and trainval.txt, which, respectively, represented the training set, validation set, test set, and summary of the training and validation sets required by the model. The numbers of the photographs in each of the four text

files' respective sets, with each image number on a distinct line, made up their contents. To ensure the generalizability of the model, the image numbers were created at random.

The function of Annocations file was mainly to store the annotation information corresponding to the leaf image. In order to train the model, a large number of data annotations of the data set must be performed; this work used Labelme as the data labeling software. The annotation file is initially stored in.json format, and then changed to a tag image in.png format by batch converting the file, as shown in Figure 3.

Setting up the testing environment and parameters

Intel Core i7-9700, 32 GB of RAM, and an Nvidia GeForce RTX 2080Ti graphics card were the specifications of the computer's processor. Model construction, training, and prediction were performed in this deep learning environment using Tensorflow-gpu1.13.2, keras2.1.5, Windows 10, 64-bit operating system, Python 3.6.13 compiled environment, CUDA10.1 architecture, and cuDNN7.4.1 Development library. When compared to other adaptive learning rate algorithms, the Adam approach is simple to use, very computationally efficient, memory-light has a quicker

FIGURE 2
Examples of apple leaf disease image.
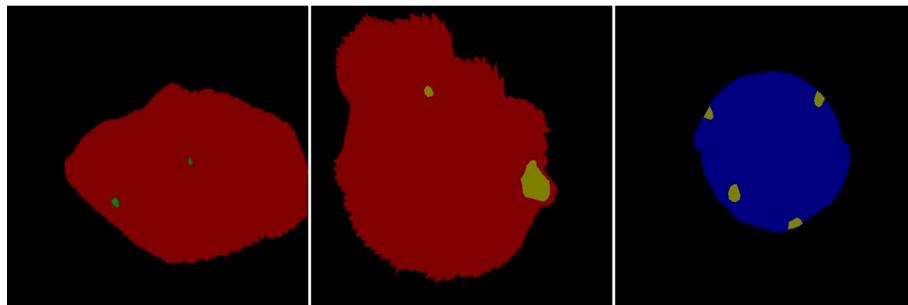
FIGURE 3
Examples image of a.png tag image.

convergence time and is invariant to diagonal gradient rescaling. In order to select the model with the best segmentation effect through interaction, this experiment trains the models using the Adam optimizer until convergence.

During training, the input image is 512*512 pixels. Padding=1 is utilized so that each input square can serve as the convolution window's center and stride=1 is used to limit the number of input parameters and processing. The output size is the same as the input when stride and padding are both set to 1. In order to nonlinearly transform the input, the activation function needs to be introduced. The activation function used in this paper is sigmoid. The whole training is divided into two stages, the freezing stage and the thawing stage. The quantity of images entered into the network at once during training is referred to as the batch size. The model training generation is known as an epoch. It can be regarded as a suitable training generation when there is a minimal difference in error between the training set and the test set. In order to ensure that the model achieves the best effect in terms of accuracy and training time, this paper sets the training generation to 200. According to the graphics performance of the operating system and the size of the image, the first 50 stages are the freezing stage, the batch size is set to 4, and the last 150 stages are the thawing stage, the batch size is set to 2, to ensure that the model achieves the best effect in terms of accuracy and training time, and avoids insufficient memory. The average value of the updated network weight in the algorithm is the initial learning rate. The maximum learning rate is set at 0.0001 in order to speed up the model training's transition into a stable learning state. The learning rate is reduced by the cosine annealing attenuation method. Period = 5 is set during training to attenuate the model once every 5 epochs and preserve it, preventing the loss of the training model in the event of a power outage or an abnormal exit during long-term training.

Model evaluation indicators

This study evaluated the classification accuracy of the model for the disease classification problem using true positive (TP, the number of times the model accurately predicts the disease type), true negative (TN, the number of times the model accurately predicts the leaf area), false positive (FP, the possibility of misjudging the leaf area as the spot area), and false negative (FN, the possibility of misjudging the spot area as the leaf area).

After the establishment of the model, it was necessary to evaluate its effectiveness. This work suggested using the mean intersection ratio MIoU (Shoaib et al., 2022), the category average pixel accuracy MPA, the precision rate Precision, and the comprehensive evaluation index F1 Score (Shoaib et al., 2022) as the evaluation index of the segmentation results in order to quantify and assess the model's performance.

To facilitate the interpretation of the evaluation metric formulas, it is assumed that the data set has a total of $k + 1$ categories. p_{ij} denotes the number of pixels for which category i is predicted to category j, p_{ii} denotes the number of pixels that are correctly predicted, and p_{ij} and p_{ji} denote the number of false negative and false positive pixels, respectively.

(1) MIoU

The average of the ratio between the intersection and concatenation of the set of pixels whose true value is the spot and the set of pixels whose predicted value is the spot is determined, as indicated in equation (4). The higher the MIoU value, the higher the overlapping degree between the projected spot area and the actual spot area.

$$\text{MIoU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \quad (4)$$

(2) MPA

Equation (5) demonstrates that MPA is the average of the percentage of total pixels that fall into the proper prediction category.

$$\text{MPA} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{P_{ii}}{\sum_{j=0}^{k} P_{ij}} \quad (5)$$

(3) Precision

The accuracy rate is defined as the proportion of actual diseased pixels to those predicted as such by the model, as indicated in equation (6). Less false detection areas are seen in the prediction results as the value increases.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

(4) Recall

The recall rate, also known as the check-all rate, is the proportion of spots that are detected to all spots in the data set, as is evident from equation (7).

$$Recall = \frac{TP}{TP+FN} \qquad (7)$$

(5) F1 Score

The F1 Score metric combines the Precision and Recall outputs, as given in equation (8). F1 Score accepts values between 0 and 1. The model's best output is represented by 1, while its worst output is represented by 0. The more correctly recognized spot pixels, the more accurate the segmentation result.

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} \qquad (8)$$

Test results and analysis

Test process

After 250 epochs of training the U-Net model, the Loss finally converged to 0.022. Figure 4 depicts the Loss's evolution throughout training epochs. It is clear from the figure that the Loss stopped dropping and stabilized around 200 epochs, indicating that the model had progressively converged at that point. The U-Net model with 200 and 250 epochs of training was chosen to compare the test results in order to determine the best model for this experiment. The findings are displayed in Table 1.

Table 1 shows that as training epochs increased, MIoU, MPA, and Precision values fell at training 250 epochs, indicating the occurrence of an overfitting phenomenon. As a result, the model in this research was chosen for training 200 epochs.

In the experiment to segment unhealthy spots, the target pixel points can be separated into two primary categories: diseased spots and healthy parts. Since the background does not include any diseased spots, it is likewise segmented into healthy parts. Three

loss functions—CEL, CEDL, and DFL—are employed in this study's ablation experiments, with U-Net serving as the main body. The experiment assesses the effectiveness of the loss functions using the loss rate and accuracy of the validation set. Table 2 presents the outcomes.

Table 2 compares the performance of the original U-Net and the improved U-Net deep learning designs using various loss functions. Verification loss, accuracy, MIoU, and MPA are employed as evaluation indicators for these variables. As can be observed, under the presumption that picture segmentation accuracy is guaranteed, the outcomes of the four parameters are 0.008, 98.86%, 91.07%, and 95.58%, respectively, after adding the DFL mixed loss function. The modified model's average loss rate dropped from 0.063 to 0.008; the lower the loss, the more accurate the model, the MPA increased by 1.06%, the prediction category correctness increased, the MIoU score rose by 1.96%, and the more the predicted illness area overlapped with the actual disease region. The challenge of distinguishing apple Alternaria blotch disease and brown spot disease with high similarity in the early stage of disease is resolved by the addition of the DFL mixed loss function, which also addresses the issue of poor segmentation performance of smaller disease points. Additionally, it lessens the disparity between simple and difficult training examples as well as the disparity between positive and negative training examples. The process by which the effective loss value of the U-Net model changes when different loss functions are applied is shown in Figure 5. The outcomes demonstrate that the DFL mixed loss function employed in this study has the smallest loss value, the fastest decline rate, and the smoothest training procedure.

The model is trained by adding various attention mechanism modules using the same experimental setting and training parameters as U-Net combined with hybrid loss function DFL. The experimental findings are displayed in Table 3 to compare the various types of segmentation MIoU.

As shown in Table 3, the accuracy of disease identification can be increased by adding SENet or ECANet, but the addition of the CBAM attention mechanism results in superior disease identification. The comparison shows that the MIoU value of smaller Alternaria blotch disease spots increases by 2.97%, indicating that the CBAM attention mechanism can effectively focus on the disease spots in the image and suppress the interference information. To address the issue of the DFL-UNet model's poor segmentation performance of smaller spots, we decided to integrate the CBAM module in this study.
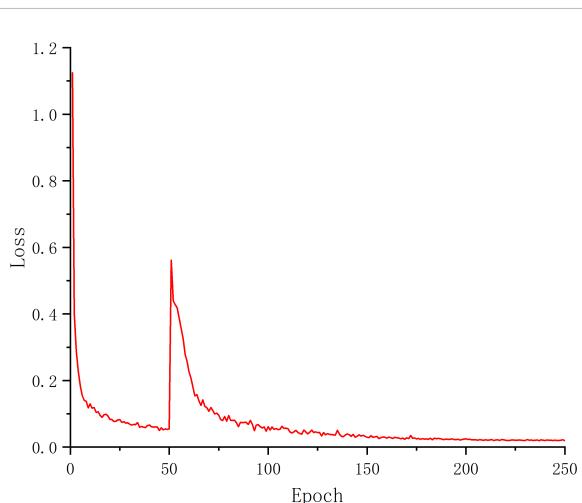

FIGURE 4
Loss curve.

TABLE 1 Comparison of segmentation results for different epochs of training.

Epoch	MIoU/(%)	MPA/(%)	Precision/(%)
200	89.11	94.52	93.53
250	88.96	94.30	93.24

TABLE 2 Experimental results of loss function ablation.

Loss function	Network	Val-acc/(%)	Val-loss	MIoU/(%)	MPA/(%)
CEL	Original U-Net	98.34	0.063	89.11	94.52
CEDL	Original U-Net	98.37	0.055	89.33	94.85
DFL	Original U-Net	98.40	0.010	90.09	95.14
CEL	Improved U-Net	98.51	0.039	90.89	95.07
CEDL	Improved U-Net	98.76	0.045	89.96	94.90
DFL	Improved U-Net	98.86	0.008	91.07	95.58

Results analysis

The segmentation performance of the Deeplabv3+ model, PSPNet model, original U-Net model, and DFL-UNet+CBAM model was compared, and the results are shown in Table 4. In this paper, MIoU, MPA, and F1 Score were all used as evaluation metrics for segmentation results under the same research object and the same experimental conditions.

As can be seen in Table 4, when comparing the four models, the MIoU, MPA, and F1 Score of the DFL-UNet+CBAM model proposed in this paper are the highest, increasing by 1.96% in MIoU value, 1.06% in MPA value, and 1.14% in F1 Score when compared with the original U-Net model. This shows that the model in this paper correctly identifies the most diseased pixels and can effectively optimize the segmentation results and obtain more. The change in MIoU value during model training is depicted in Figure 6, and it is also obvious from the change curve that the model used in this paper has the greatest MIoU value, suggesting the highest overlap between the predicted spot area and the actual spot area.

Table 5 compares the segmentation performance of smaller spots before and after model modification using MIOU, MPA, Precision, Recall, and F1 scores as assessment metrics. This comparison is done to indicate the benefit of the suggested method in recognizing smaller spots.


FIGURE 5
Loss curves of different loss functions.

As can be seen from Table 5, compared to the original U-Net model, the segmentation of Alternaria blotch disease, the MIoU value increased by 4.41%, the MPA value increased by 4.13%, the Precision increased by 1.49%, the Recall increased by 4.13%, and the F1 Score increased by 2.81%; in the segmentation of brown spots, MIoU values increased by 1.18%, MPA values by 0.6%, Precision by 0.78%, Recall by 0.6%, and F1 Score by 0.69%. The spot diameter of the Alternaria blotch disease is 0.2-0.3cm in the early stage, 0.5-0.6cm in the middle and late stages, and the spot diameter of the brown spot disease is 0.3-3cm. Obviously, brown spot spots are larger than Alternaria blotch spots. The segmentation performance of smaller disease spots has increased more noticeably, according to the quantitative analysis results, proving that the model's capacity to segment smaller disease spots has greatly improved.

Additionally, the proposed model's training and validation performance are assessed using the training set F1 score, validation set F1 score, a training set loss, and validation set loss. This is done to further validate the performance of the model segmentation. The loss value is used to quantify the discrepancy between the model's true value and its predicted value, and the F1 score is calculated as a weighted average of Precision and Recall metrics. Better model robustness is associated with smaller loss functions. The training score determines the generalization ability of the algorithm in its training samples. The verification score determines the optimal model (Srinivasu et al., 2022). Figure 7 displays the model's performance in relation to the hyperparameters.

In this study, we used a trained semantic segmentation model to predict apple leaf disease in laboratory and field environments. The image dataset must meet two criteria: first, it must allow for the simultaneous occurrence of various illnesses on the same leaf; and second, it must allow for the presence of complicated backgrounds in some images to guarantee the data images' excellent generalization ability.

In comparison to the Deeplabv3+ model, the PSPNet model, and the original U-Net model, the segmentation results of the DFL-UNet+CBAM model utilized in this paper are shown in Figure 8 for the test set of apple disease leaf photos.

The prediction outcomes of single-leaf spot segmentation against various backgrounds are shown in Figure 8. Figure 8 shows it abundantly clear that the network structure suggested in this paper achieves more accurate segmentation for apple leaf spots and produces better segmentation results for both the disease location on the leaf and the size of the spot area. This network structure is also more accurate than other networks used in this
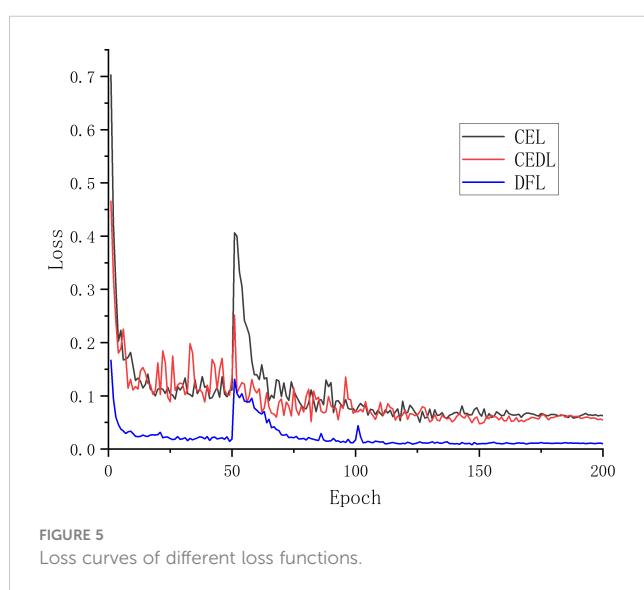
TABLE 3　Comparison of the cross-merge ratio (MIoU) for each category of the model after adding the attention mechanism.

Segmentation Model	leaf/(%)	Alternaria blotch/(%)	Brown spot disease/(%)
DFL-UNet	93.70	76.05	84.12
DFL-UNet+SENet	94.05	78.56	84.93
DFL-UNet+ECANet	93.43	78.99	85.11
DFL-UNet+CBAM	94.86	79.02	85.28

paper. When recognizing brown spot disease, the Deeplabv3+ model in Figure 8C incorrectly recognized the green halo area surrounding the illness spot as the disease spot; the PSPNet model in Figure 8D has a condition where the object boundary segmentation is discontinuous and the segmentation result is rough, the border between the leaf and the backdrop is hazy, and there is a missing section for the area affected by the brown spot disease. Analysis of the segmentation results of the model proposed in this paper demonstrates that the model in this paper can segment the semantic objects completely, finely, and accurately, and it is apparent from Figure 8F that the recognition results of the disease spots and the segmentation results of the edges of the disease spots in this paper are more accurate.

Comparing Figure 8C and Figure 8D, it can be seen that the network structure of the proposed model performs well in the segmentation of smaller spots. Although the U-Net model in Figure 8E identified smaller spots in the apple Alternaria leaf spot and brown spot categories of foliar diseases, the identified spot area was incomplete. In contrast, the model in Figure 8F accurately identified the smaller spots in the categories of apple ringspot and brown spot, and the recognition results were more accurate.

Discussion

Semantic segmentation and attention mechanisms have been widely used in the realm of disease recognition. An ASPP (Atrous Spatial Pyramid Pooling)-based DeepLabV3+ semantic segmentation network model, for instance, was developed by Li L et al. (Li et al., 2023). The experimental findings revealed that the model's average pixel accuracy (MPA) and average intersection (MIoU) reached 97.26% and 83.85%, respectively. Additionally, Li Q et al. (Li et al., 2021) proposed an integrated U-Net segmentation model for small sample datasets, merging U-edge Net's features and high-level features using ASPP. The experimental findings

demonstrated that the method significantly increased the segmentation accuracy of the target fruits as well as the model's capacity for generalization.

The segmentation task of apple leaves and spot areas was carried out in this study using three traditional semantic segmentation network models (DeepLabV3+, PSPNet, and U-Net). The segmentation performance of the model was evaluated throughout the experiment. Also, the performance of the model is addressed in relation to the implications of various loss functions and attention mechanisms. Following are our findings:

1. Three semantic segmentation network models (DeepLabV3 +, PSPNet, and PSPNet) were compared and their segmentation and convergence capabilities for the apple leaf and speckle regions were examined. The findings indicate that PSPNet and Deeplabv3+ are not as effective in segmenting data as the U-Net network model.

2. Investigated is how the U-Net network model chooses its loss function. According to the results, the addition of the DFL hybrid loss function improves the segmentation performance and classification capacity of the model. The average loss rate val-loss lowers from 0.063 to 0.008, the MIoU index increases by 1.96%, and the MPA increases by 1.06%.

3. Compare the different U-Net attention mechanism modules. The findings demonstrate that the addition of the CBAM attention mechanism improves the disease recognition effect. Comparatively, it is discovered that the MIoU value of the smaller speckle leaf spot disease spot is increased by 2.97%, demonstrating that the CBAM attention mechanism can concentrate on and pay attention to the disease spot in the image, as well as effectively suppress the interference information, which enhances the model's focus on the target channel and spatial information.

TABLE 4　Comparison table of segmentation performance of different models.

Model	MIoU/(%)	MPA/(%)	F1 Score/(%)
Deeplabv3+	85.94	92.04	91.80
PSPNet	83.49	86.81	90.40
U-Net	89.11	94.52	94.02
DFL-UNet+CBAM	91.07	95.58	95.16

FIGURE 6
Comparison of different model segmentation mean intersection over union.
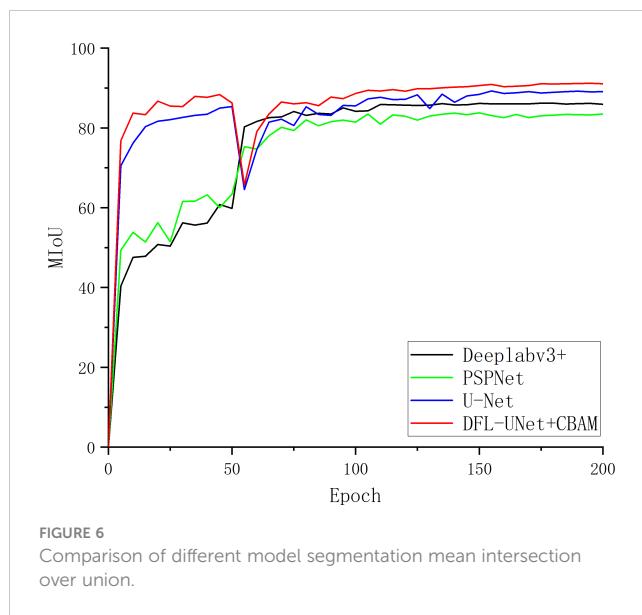
In the prior research, the loss function of the model is typically a single loss function. In this study, to enhance the segmentation performance and achieve more precise segmentation of leaves and disease spots under natural conditions, we fused two loss functions and added attention mechanisms to both the two effective feature layers extracted by the backbone network and the outcomes of the first upsampling.

Overall, our technique demonstrates good adaptability in the single background and complicated background segmentation and detection of leaf spots. But because there are so many distractions in the natural world (such as uneven lighting), incorrect detection and missed detection will always happen there. In order to test the segmentation performance of the model, Figure 9 uses the relatively smaller and more challenging-to-identify Alternaria blotch disease as an example. It then displays the segmentation prediction results of the diseased leaves and disease spots in the multi-leaf image in the natural environment. The findings demonstrate that the disease spot segmentation effect is effective when the uneven light shadow coverage is varied, however, there is a false detection part between the leaf and the background.

The target leaves' edges are difficult to extract because the background of the image in the outdoor scene has multiple leaves overlapping each other. Additionally, there are shadows in the leaf images due to uneven lighting or because of curling and folding, which makes the segmentation more challenging.

The area of light irradiation to the leaves is also diverse, as illustrated in Figure 9, due to different shooting angles and self-

curling factors. As a result, diseased leaves are concealed by other leaves or object shadows, which causes the pigment imbalance problem. Diagram 9 (a) (b) The disease leaf identification is insufficient because only a small portion of the disease leaf's edge was impacted by other leaves, a phenomenon known as missed detection; in Figure 9(c), the disease leaf edge segmentation is inaccurate because there is cross-over between leaves and a light uneven dual impact; as seen in Figure 9(d), the diseased leaf shadow is heavier Part of the incorrect check for the background area, the overall image tone is dark, the color of the measured target is distorted.

The following issues still need to be resolved even though we explored the segmentation recognition of smaller spots in apple leaves in this work and increased the segmentation effect and recognition accuracy of smaller spots.

1. The ability to quickly diagnose diseases in fruit trees is crucial for practical production, so future research should focus on enhancing the network structure to reduce the model segmentation time. This will help fruit farmers quickly confirm the diagnosis of diseases in fruit trees and quickly apply pesticides.
2. In actual, there are frequently several leaves in a single image and the leaves are set against a complicated background. The presence of disease spots on many leaves is not taken into account in this work. Consequently, to enhance the segmentation performance of disease leaves and thereby enhance the precision of disease spot recognition, the model needs to be further enhanced in the upcoming research.
3. The actual development of disease species is complex and varied. Despite the fact that the method described in this paper enhances the segmentation performance of smaller spots in apple leaf diseases and the recognition precision of difficult-to-classify diseases, the disease species in the training data set still need to be increased, and the disease species can be increased later to improve the recognition and segmentation ability of the model for various diseases and make the model broadly applicable.

Conclusion

In practice, the naked eye can easily misinterpret the type of disease and thus overuse pesticides, which in turn affects apple

TABLE 5 Analysis of quantitative results of U-Net and improved U-Net.

| Model	Disease types	MIoU/(%)	MPA/(%)	Precision/(%)	Recall/(%)	F1 Score/(%)
Original U-Net	Alternaria blotch	74.61	84.93	86.02	84.93	85.47
	Brown spot	84.1	92.7	90.06	92.7	91.36
Improved U-Net	Alternaria blotch	79.02	89.06	87.51	89.06	88.28
	Brown spot	85.28	93.3	90.84	93.3	92.05
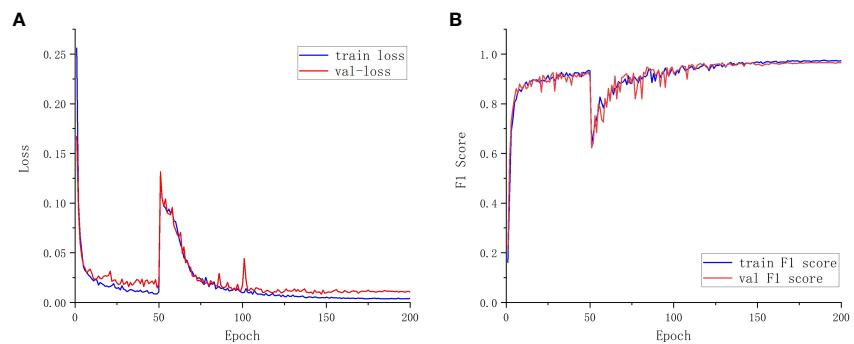
FIGURE 7

Training and validation set details. **(A)** Loss function curve. **(B)** F1 score curve.

production. Therefore, disease diagnosis must be easier, faster, and more accurate, while the type of disease must be analyzed and determined. Apple leaf spot is very small and has similar characteristics when the disease first appears, while the actual orchard environment has different light conditions, overlapping leaf shade, etc. A deep learning-based apple leaf disease spot segmentation technique is suggested for apple leaf disease recognition by utilizing CNN's strong feature extraction capabilities in order to minimize the influence on disease spot segmentation. The core network architecture used by the method is a convolutional neural network called U-Net, and to better extract picture features, its structure and parameters have been modified and optimized. The identification of apple leaf disease depends directly on the precision of the segmentation method. In order to address the issues of low recognition accuracy and subpar performance of smaller spot segmentation in apple leaf disease recognition, this paper uses apple leaf Alternaria blotch and brown

spot as its research object. It then proposes a method of spot segmentation and disease recognition based on hybrid loss function and CBAM. The following conclusions were obtained from the study:

1. To deal with the issue of poor performance in segmenting smaller spots in apple leaves, a model for apple leaf disease segmentation based on hybrid loss function and CBAM network has been developed. Firstly, the model using mixed loss function of Dice Loss and Focal Loss has swapped out the original cross entropy function, which has given larger weight to the samples that are difficult to classify, making the model pay more attention to the target with smaller pixel proportion. Secondly, the backbone network's two useful feature layers and the outcomes of the first upsampling have been combined with the CBAM module to complete the extraction of pixel features and disease spot
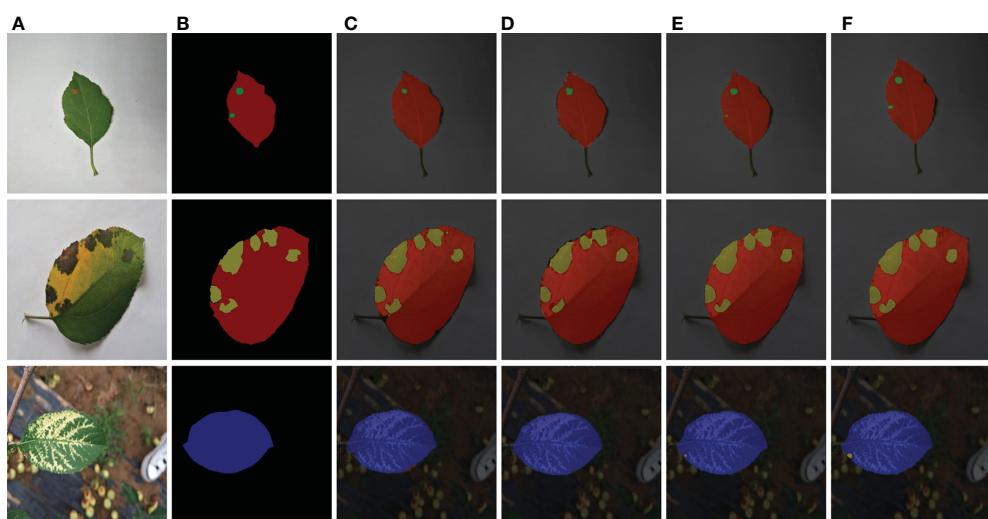

FIGURE 8

Comparison of segmentation results of various models. **(A)** Original images. **(B)** Ground truth. **(C)** Deeplabv3+ segmentation results. **(D)** PSPNet segmentation results. **(E)** U-Net segmentation results. **(F)** Improved U-Net segmentation results.
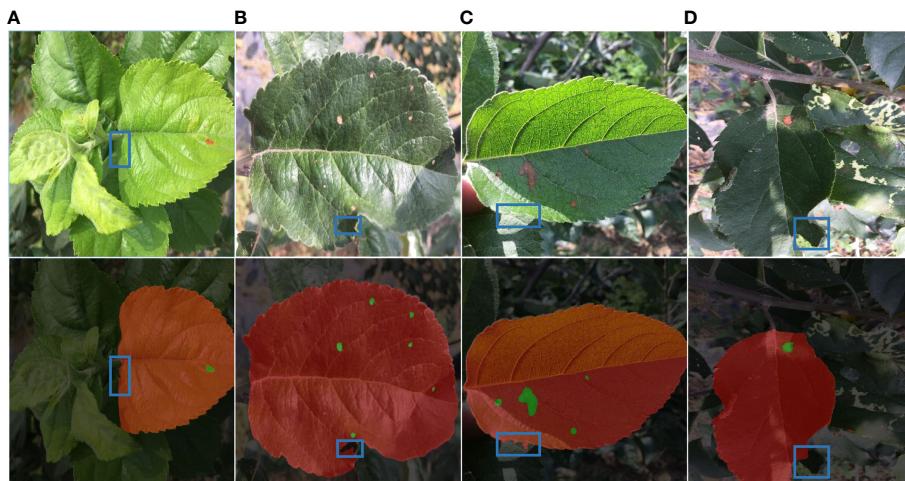
FIGURE 9
Detection fault analysis. **(A)** Leaf occlusion. **(B)** Self-crimp factor. **(C)** Leaf folding. **(D)** Insufficient light.

segmentation for apple Alternaria blotch and brown spot. This has caused the model to pay more attention to the regions with important information.

2. MIoU values in DFL-UNet+CBAM model employed in this study were 91.07%, MPA values o were 95.58%, and F1 Score values were 95.16%. These values were higher than those of the original U-Net model by 1.96%, 1.06%, and 1.14% respectively, and the illness identification impact was also enhanced. The segmentation result images have also shown that the DFL-UNet+CBAM model has had better segmentation and recognition capabilities, can more precisely identify smaller disease spot areas, improves the detection and recognition accuracy of smaller disease spots, better satisfies the requirements of apple leaf disease recognition, and provides a basis for the diagnosis of apple leaf diseases.

3. In the multi-blade environment of nature, several leaves may coexist on a single map, and various illnesses may be present on the leaves. The experimental results demonstrate that the semantic segmentation model of apple leaf diseases trained in this paper using a single leaf dataset can not only detect a single background in the laboratory but can also be used to detect apple leaf diseases in the complex background of the natural environment; it can not only detect single objects of single and multiple leaves, but it can also detect multiple objects of single leaves, demonstrating powerful segmentation performance.

Research demonstrates that the model can ensure segmentation accuracy in complicated orchard environments as well as laboratores, particularly when it comes to the edge segmentation accuracy of smaller disease spots. The suggested method performs segmentation better than other methods, and the model has good generalizability. In the future, it might serve as a technical foundation for the segmentation, categorization, and precise management of plant leaf disease spots.

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

XZ drafted the original draft. XZ and DL reviewed and revised the original draft. XLiu provided guidance. TS and XLin collected the data. ZR provided financial supports and retouched the manuscript. All authors contributed to the article and approved the submitted version.

Funding

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

Ayaz, H., Rodríguez-Esparza, E., Ahmad, M., Oliva, D., Pérez-Cisneros, M., and Sarkar, R. (2021). Classification of apple disease based on non-linear deep features. *Appl. Sci. (Switzerland)* 11, 1–13. doi: 10.3390/app11146422

Chen, L. C., Papandreou, G., Kokkinos, I., and Murphy, K. (2018a). DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 834–848. doi: 10.1109/TPAMI.2017.2699184

Chen, G., and Qin, H. (2022). Class-discriminative focal loss for extreme imbalanced multiclass object detection towards autonomous driving. *Vis. Comput.* 38, 1051–1063. doi: 10.1007/s00371-021-02067-9

Geetharamani, G., and Pandian, J. A. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Comput. Electrical Eng.* 76, 323–338. doi: 10.1016/j.compeleceng.2019.04.011

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2020). Squeeze-and-Excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023. doi: 10.1109/TPAMI.2019.2913372

Hussein, B. R., Malik, O. A., Ong, W. H., and Slik, J. W. F. (2021). Automated extraction of phenotypic leaf traits of individual intact herbarium leaves from herbarium specimen images using deep learning based semantic segmentation. *Sensors* 21, 1–23. doi: 10.3390/s21134549

Jain, P. K., Dubey, A., Saba, L., Khanna, N. N., Laird, J. R., Nicolaides, A., et al. (2022). Attention-based UNet deep learning model for plaque segmentation in carotid ultrasound for stroke risk stratification: an artificial intelligence paradigm. *J. Cardiovasc. Dev. Dis.* 9, 1–30. doi: 10.3390/jcdd9100326

Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., and Menaka, R. (2020). Attention embedded residual CNN for disease detection in tomato leaves. *Appl. Soft Computing J.* 86, 1–27. doi: 10.1016/j.asoc.2019.105933

Khan, A. I., Quadri, S. M. K., Banday, S., and Latief Shah, J. (2022). Deep diagnosis: a real-time apple leaf disease detection system based on deep learning. *Comput. Electron Agric.* 198, 1–12. doi: 10.1016/j.compag.2022.107093

Li, Q., Jia, W., Sun, M., Hou, S., and Zheng, Y. (2021). A novel green apple segmentation algorithm based on ensemble U-net under complex orchard environment. *Comput. Electron. Agric.* 180, 1–10. doi: 10.1016/j.compag.2020.105900

Li, L., Wang, B., Li, Y., and Yang, H. (2023). Diagnosis and mobile application of apple leaf disease degree based on a small-sample dataset. *Plants* 12, 1–15. doi: 10.3390/plants12040786

Li, X., Zhou, Y., Liu, J., Wang, L., Zhang, J., and Fan, X. (2022). The detection method of potato foliage diseases in complex background based on instance segmentation and semantic segmentation. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.899754

Lin, K., Gong, L., Huang, Y., Liu, C., and Pan, J. (2019). Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00155

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2020). Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 318–327. doi: 10.1109/TPAMI.2018.2858826

Liu, B. Y., Fan, K. J., Su, W. H., and Peng, Y. (2022). Two-stage convolutional neural networks for diagnosing the severity of alternaria leaf blotch disease of the apple tree. *Remote Sens (Basel)* 14, 1–17. doi: 10.3390/rs14112519

Liu, B., Zhang, Y., He, D. J., and Li, Y. (2018). Identification of apple leaf diseases based on deep convolutional neural networks. *Symmetry (Basel)* 10, 1–16. doi: 10.3390/sym10010011

Ma, R., Wang, J., Zhao, W., Guo, H., Dai, D., Yun, Y., et al. (2022). Identification of maize seed varieties using MobileNetV2 with improved attention mechanism CBAM. *Agriculture* 13, 11. doi: 10.3390/agriculture13010011

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

Ngugi, L. C., Abelwahab, M., and Abo-Zahhad, M. (2020). Tomato leaf segmentation algorithms for mobile phone applications using deep learning. *Comput. Electron Agric.* 178, 1–15. doi: 10.1016/j.compag.2020.105788

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

Santos, T. T., de Souza, L. L., dos Santos, A. A., and Avila, S. (2020). Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *Comput. Electron Agric.* 170, 1–17. doi: 10.1016/j.compag.2020.105247

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. doi: 10.1109/TPAMI.2016.2572683

Shoaib, M., Hussain, T., Shah, B., Ullah, I., Shah, S. M., Ali, F., et al. (2022). Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1031748

Srinivasu, P. N., Shafi, J., Krishna, T. B., Sujatha, C. N., Praveen, S. P., and Ijaz, M. F. (2022). Using recurrent neural networks for predicting type-2 diabetes from genomic and tabular data. *Diagnostics* 12, 1–30. doi: 10.3390/diagnostics12123067

Tassis, L. M., Tozzi de Souza, J. E., and Krohling, R. A. (2021). A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images. *Comput. Electron Agric.* 186, 1–12. doi: 10.1016/j.compag.2021.106191

Too, E. C., Yujian, L., Njuki, S., and Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron Agric.* 161, 272–279. doi: 10.1016/j.compag.2018.03.032

Wang, P., Niu, T., Mao, Y., Zhang, Z., Liu, B., and He, D. (2021). Identification of apple leaf diseases by improved deep convolutional neural networks with an attention mechanism. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.723294

Wang, L., Wang, C., Sun, Z., and Chen, S. (2020). An improved dice loss for pneumothorax segmentation by mining the information of negative areas. *IEEE Access* 8, 167939–167949. doi: 10.1109/ACCESS.2020.3020475

Yu, H. L., Cheng, X. H., Li, Z. Q., Cai, Q., and Bi, C. G. (2022). Disease recognition of apple leaf using lightweight multi-scale network with ECANet. *CMES-COMPUTER MODELING IN Eng. Sci.* 132, 711–738. doi: 10.32604/cmes.2022.020263

Zhong, Y., and Zhao, M. (2020). Research on deep learning in apple leaf disease recognition. *Comput. Electron Agric.* 168, 1–6. doi: 10.1016/j.compag.2019.105146

A latent scale model to minimize subjectivity in the analysis of visual rating data for the National Turfgrass Evaluation Program

Yuanshuo Qu[1]*, Len Kne[2], Steve Graham[3], Eric Watkins[4]
and Kevin Morris[1]

[1]National Turfgrass Evaluation Program, Beltsville, MD, United States, [2]U-Spatial, University of
Minnesota, Minneapolis, MN, United States, [3]U-Spatial, University of Minnesota, Duluth, MN, United
States, [4]Department of Horticultural Science, University of Minnesota, St. Paul, MN, United States

Introduction: Traditional evaluation procedure in National Turfgrass Evaluation Program (NTEP) relies on visually assessing replicated turf plots at multiple testing locations. This process yields ordinal data; however, statistical models that falsely assume these to be interval or ratio data have almost exclusively been applied in the subsequent analysis. This practice raises concerns about procedural subjectivity, preventing objective comparisons of cultivars across different test locations. It may also lead to serious errors, such as increased false alarms, failures to detect effects, and even inversions of differences among groups.

Methods: We reviewed this problem, identified sources of subjectivity, and presented a model-based approach to minimize subjectivity, allowing objective comparisons of cultivars across different locations and better monitoring of the evaluation procedure. We demonstrate how to fit the described model in a Bayesian framework with Stan, using datasets on overall turf quality ratings from the 2017 NTEP Kentucky bluegrass trials at seven testing locations.

Results: Compared with the existing method, ours allows the estimation of additional parameters, i.e., category thresholds, rating severity, and within-field spatial variations, and provides better separation of cultivar means and more realistic standard deviations.

Discussion: To implement the proposed model, additional information on rater identification, trial layout, rating date is needed. Given the model assumptions, we recommend small trials to reduce rater fatigue. For large trials, ratings can be conducted for each replication on multiple occasions instead of all at once. To minimize subjectivity, multiple raters are required. We also proposed new ideas on temporal analysis, incorporating existing knowledge of turfgrass.

KEYWORDS

NTEP, visual ratings, cultivar evaluation, subjectivity minimization, Bayesian model

1 Introduction

The National Turfgrass Evaluation Program (NTEP) is an internationally renowned turfgrass research program. Starting from 1981, NTEP has coordinated trials and collected data on a variety of turfgrass species at locations across the United States and Canada (Xie et al., 2022). At each testing location, replicated turf plots of different cultivars are established, maintained, and visually evaluated by trained raters periodically on various traits of interest. Experienced raters usually mentor new raters following rating guidelines set by NTEP. Evaluated traits have traditionally included overall quality, color, density, resistance to diseases and insects, tolerance to biotic or abiotic stresses, and more recently expanded to drought and traffic tolerance. Over the years, NTEP has created a unique data repository, providing rich information for characterizing and selecting turfgrass cultivars for various applications.

NTEP adopted a 1-9 integer scale to assess traits of selected turfgrass cultivars (hereinafter referred to as the NTEP scale). It was originally used by turfgrass researchers in the 1980s in the northeastern region of the United States (personal communication with Dr. Bill Meyer of Rutgers University), which resembles the 9-point hedonic scale. Developed by David R. Peryam and his colleagues (Peryam and Girardot, 1952; Peryam and Pilgrim, 1957), the 9-point hedonic scale was originally used to measure the food, i.e., the stimuli, preferences of soldiers, i.e., the subjects, in the U.S. Armed Forces in the 1950s. Since then, it has become the most widely used scale for testing consumer preferences and acceptability of foods and beverages (Lim et al., 2009). The original 9-point hedonic scale is a balanced bipolar scale centered around a neutral position with four positive and four negative categories on each side. The categories are labeled with phrases ranging from "Dislike Extremely" to "Like Extremely" (Table 1), representing a continuum from dislikes to likes.

Response to the 9-point hedonic scale is an ordinal variable as its categories have a natural order (Seddon et al., 2001). In subsequent analysis, the categories are generally assigned with numerical values from 1 to 9, respectively, such that parametric statistical models can be utilized. For the NTEP scale, a trained rater walks through all plots in serpentine order in each rating event, assigning an integer from 1 to 9 directly for a particular trait of

interest where 1 is typically the poorest/lowest and 9 is the best/highest. Similar to analyzing responses to a 9-point hedonic scale, the analysis of NTEP rating data treats the ordinal variables as numerical values, which may lead to serious errors, such as increased false alarms, i.e., detecting non-existing effects, failures to detect effects, and even inversions of differences among groups (Bürkner and Vuorre, 2019). There is abundant literature, e.g., Lim et al. (2009), Liddell and Kruschke (2018), on the reasons for these problems. Some important ones are summarized here.

1. The categories in the 9-point hedonic scale are not equidistant, which was first discovered by the Psychometric Laboratory at the University of Chicago (Jones and Thurstone, 1955; Jones et al., 1955), and confirmed in later studies (Moskowitz, 1971; Moskowitz and Sidel, 1971; Moskowitz, 1977; Moskowitz, 1980).
2. The 9-point hedonic scale lacks an absolute zero point. While there is a neutral position (i.e., the INDIFFERENT category or the "5"), it varies from subject to subject, even across different measurements by the same subject.
3. The general tendency of subjects to avoid using the extreme categories (Hollingworth, 1910; Stevens and Galanter, 1957; Parducci and Wedell, 1986) makes the scale vulnerable to ceiling and flooring effects. This truncates the 9-point scale, limits the scale's ability to identify extreme stimuli, and skews the response data.

As a derivation of the original 9-point hedonic scale, the NTEP scale also yields ordinal data. Such data only provide rudimentary information on the hedonic magnitude and cannot directly be used to compare hedonic perceptions across different raters. In the current evaluation process, a turf plot's rating for a specific trait, e.g., turf quality, depends on the rater's severity in the rating event. Given the same plot, it will likely score higher when the rater is lenient or lower when severe, giving rise to subjectivity. In other words, for a specific rater's turf quality ratings, we know a "3" plot has better turf quality than a "2" plot. But we cannot conclude a "3" plot rated by A is better than a "3" plot rated by B in turf quality without adjusting for rater severity. Considering the temporal nature of the evaluation process, even for the same rater on the

TABLE 1 Replication of the questionnaire designed for studying soldiers' preferences in the field.

	FOOD ITEM	LIKE				INDIFFERENT	DISLIKE			
Not Tried	Cream Gravy	Like Extremely	Like Very Much	Like Moderately	Like Slightly	Neither Like Nor Dislike	Dislike Slightly	Dislike Moderately	Dislike Very Much	Dislike Extremely
Not Tried	Bread Putting	Like Extremely	Like Very Much	Like Moderately	Like Slightly	Neither Like Nor Dislike	Dislike Slightly	Dislike Moderately	Dislike Very Much	Dislike Extremely
Not Tried	Cheese	Like Extremely	Like Very Much	Like Moderately	Like Slightly	Neither Like Nor Dislike	Dislike Slightly	Dislike Moderately	Dislike Very Much	Dislike Extremely
Not Tried	French Fried Onions	Like Extremely	Like Very Much	Like Moderately	Like Slightly	Neither Like Nor Dislike	Dislike Slightly	Dislike Moderately	Dislike Very Much	Dislike Extremely
Not Tried	Lettuce Wedges	Like Extremely	Like Very Much	Like Moderately	Like Slightly	Neither Like Nor Dislike	Dislike Slightly	Dislike Moderately	Dislike Very Much	Dislike Extremely

same trait, consistency is not guaranteed at different times of the year. Another source of subjectivity relates to the scale categories, which are not equal distances or of the same levels. To meaningfully aggregate data collected from different rating events across different testing sites, both sources of subjectivity need to be addressed. However, current methods, e.g., the additive main effect and multiplicative interaction (AMMI) method, analysis of variance (ANOVA) (Ebdon and Gauch Jr., 2002a; Ebdon and Gauch Jr., 2002b), and linear mixed model (LMM), are not adequate and suffer the same errors when they were applied to ordinal data directly. Inspired by Rasch Rating Scale Model (Andrich, 1978), we propose a latent scale model to minimize subjectivity, hereinafter referred to as NTEP RSM (NTEP Rating Scale Model), allowing more objective comparisons of cultivars across different raters and research groups. We also demonstrate how to fit the described model in a Bayesian framework, using datasets on overall turf quality ratings in the 2017 NTEP Kentucky bluegrass trials. The model is programmed in Stan (Lee et al., 2017) *via* Python. Stan is a probabilistic programming language for statistical modeling, inference, and computation. Although demonstrations are done for overall turf quality rating, this approach works for other traits of interest evaluated using the 1-9 NTEP rating scale.

2 Model specifications

2.1 NTEP RSM

We started by constructing a latent scale based on the probability distribution of raw ordinal data. The model predicts the decision between two adjacent categories using a threshold parameter on the latent scale. The 1-9 scale is re-indexed in the following sections as 0-8 categories for conciseness in mathematical notations. At a given test location, let Y_{ni} denote the rating assigned to plot n in rating event i, the logarithmic ratio of the probability of plot n assigned to category s to that of plot n assigned to $s-1$ can be expressed by the following equation,

$$ln\left[\frac{Pr(Y_{ni} = s)}{Pr(Y_{ni} = s - 1)}\right] = \theta_n - \beta_i - \tau_s \qquad (1)$$

where

$i=1,2,\ldots,I$ is the index for each rating event during the trial;

$n=1,2,\ldots,N$ is the index for each plot;

$s=1,2,\ldots,M$ is the index for category thresholds;

$M(M \le 8)$ is both the maximum rating score after reindexing and the number of thresholds;

θ_n is the perceived turf quality of plot n in a specific rating event;

β_i measures rating severity in rating event i;

τ_s is the threshold at which at $Pr(Y=s-1) = Pr(Y=s)$.

Constraints were placed on β_I and τ_S to add a meaningful zero to the scale. Both parameters were constrained to be the negative sum of the other parameters, respectively. We further assume θ, β,
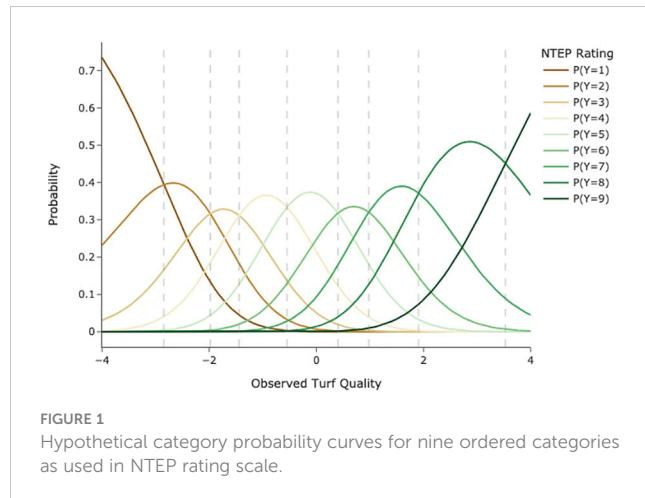
and $\boldsymbol{\tau}$ are normally distributed. For an unbiased rater in a rating event ($\beta=0$), the probability density curves for each category are illustrated in Figure 1. The vertical dash lines indicate category thresholds located at the points where the probability of a cultivar being assigned to two adjacent categories is equal. Note that these thresholds are not necessarily equidistant. In Figure 1, if a cultivar is located in a category (i.e., between two adjacent thresholds), then the response in that category has the greatest probability. The x-axis represents the constructed latent scale. It is continuous and equidistant, with a zero indicating the average level of overall turf quality. While the average level in individual rating events might vary ($\beta\neq0$), we assume the average levels for each research group at different test locations are the same, allowing scale matching across different testing locations. Once subjectivity effects, i.e., $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, were estimated and removed, $\boldsymbol{\theta}$ can be further analyzed. In this study, we partitioned $\boldsymbol{\theta}$ into cultivar and plot location effects, that is,

$$\theta = \eta + \mathcal{LOC} \qquad (2)$$

where $\boldsymbol{\eta}$ is the cultivar effect, reflecting the intrinsic quality of a cultivar, and $\mathcal{LOC}$ is the plot location effect due to spatial heterogeneity of the field. We further assume cultivar effects follow normal distributions with a mean of 0 and a variance of σ^2. The plot location effect was modeled as a Gaussian process with a zero mean and covariance function K,

$$\mathcal{LOC}(\cdot) \sim N(\mathbf{0}, K(\cdot)) \qquad (3)$$

The covariance function $K(\cdot)$ implemented here is an exponential quadratic function. For two plots i and j in the same trial at a specific testing location,

$$K(\cdot \,|\alpha, \rho, \sigma_e)_{ij} = \alpha^2 \exp\left(\frac{d_{ij}^2}{2\rho^2}\right) + \delta_{ij}\sigma_e^2 \qquad (4)$$

where α, ρ, and σ_e are hyperparameters defining the covariance function; δ_{ij} is the Kronecker delta function with value 1 if $i = j$ and 0 otherwise; d_{ij} is the Euclidean distance between centers of the two plots. As this is a Bayesian model, priors for parameters and hyperparameters are required. We adopted weakly informative priors: $t_3(0,1)$ for α, σ and σ_e; Inv–Gamma(5,5) for ρ.

2.2 Parameter recovery with NTEP RSM

To ensure that model parameters are identifiable, the following parameter recovery test was performed to evaluate the model. We first generated a synthetic dataset from 3 replications of 10 cultivars rated monthly for 5 years by 5 raters. The entry effects are random draws from a normal distribution with a mean of 0 and a standard deviation of 0.7 (σ = 0.7). Plot location effects are generated from a Gaussian process with an assigned mean vector and covariance matrix with α = 0.15, ρ = 2.5, σ_e = 0.2. Rating severity is a vector of five evenly spaced numbers over [−0.8,0.8], and category threshold is a vector of eight evenly spaced numbers over [−2,2]. All parameters, functions, and simulated data can be found in the Github repository. The simulated data were fit to the NTEP RSM for parameter recovery.

2.3 Linear mixed model

To compare with the existing method, we also implemented the following LMM for each testing location,

$$\mathbf{Y} = \boldsymbol{\eta} + \mathbf{u} + \epsilon \tag{5}$$

in which quality rating, $\mathbf{Y}$, was treated as a continuous variable and partitioned into a fixed effect of cultivars, $\boldsymbol{\eta}$, and a random effect of rating event, $\boldsymbol{u}$. ϵ denotes the residual that the model does not explain.

2.4 Model implementation

The NTEP RSM model is implemented in Stan (version 2.29.1) with a Python interface (version 3.10.4). The same model was fitted to data collected from each trial location, and posterior sampling of model parameters was generated by four Markov chain Monte Carlo chains, each with 1,000 iterations. The first 500 iterations were discarded to minimize the effect of initial values, and the rest were thinned by taking every other sample to reduce sample autocorrelation. The convergence of chains was confirmed *via* visual inspection and examining the $\hat{R}$ values of all parameters and the log posteriors. Model codes and output files can be found at https://github.com/QhenryQ/ntep-rsm. The LMM is implemented with the Python package Statsmodels (Seabold and Perktold, 2010).

3 Results and discussions

3.1 Preliminary data analysis

Kentucky bluegrass is a cool-season turfgrass that grows best when temperatures are between 60-75°F and goes dormant in hot, dry summer and cold winter. Given this behavior, turf quality data is only collected from May to October in northern trial locations, while in the southern trial locations, data is usually collected all year round. Figure 2 presents monthly histograms for all the raw turf quality rating data. In most months, the quality rating showed good symmetry and central tendency around 5 or 6. In January and February, turf quality ratings were only available from Raleigh, NC, and Stillwater, OK. We noticed decreased turf quality ratings and the number of categories assigned in both locations. For example, the February overall turf quality ratings at Stillwater, OK, were found to have a range of [3, 6], with a median of 4. This is presumably due to raters' adjustment to the dormancy of Kentucky bluegrass. The significant reduction of turf quality in dormancy makes it difficult for raters to distinguish cultivars. Ceiling and flooring effects were also observed at other locations, e.g., the overall turf quality data at East Lansing, MI, and Raleigh, NC, ranged from 2 to 9, while that for data at West Lafayette, IN, from 2 to 8.
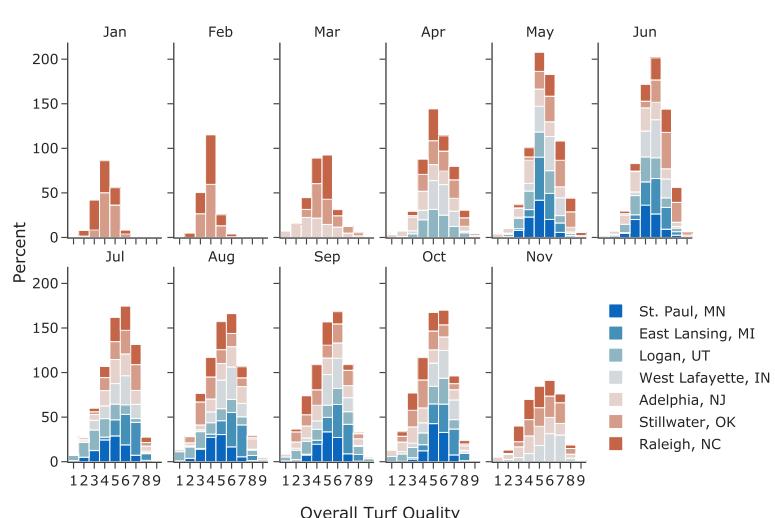

FIGURE 2
Histogram of raw overall turf quality ratings for each month at seven test locations.

3.2 NTEP RSM results

3.2.1 Category thresholds

"How is Rater A's 5 different from Rater B's 5?" This type of question is inevitable when it comes to the comparison of cultivars following the current NTEP procedure. However, such a question cannot be answered without proper definitions of categories, which in our model, are done by identifying category thresholds. These thresholds are points on the latent scale at which a rater is equally likely to select two adjacent response options (Andrich and Luo, 2003). We also assumed there are fixed distances among the category thresholds for raters within the same research group at the same location. This assumption is reasonable given that experienced raters of the same research group usually train newer raters. Estimation of category thresholds from the data provides important feedback on category definitions and how the scale is utilized by each research group, allowing us to ensure raters are adequately differentiating cultivars. When adjacent thresholds are too far apart, a category becomes too wide and less informative; on the other hand, when adjacent thresholds are close, a category becomes too narrow, indicating underutilization of the scale (see Guidelines for Rating Scales and Andrich Thresholds). We examined the non-terminal categories used at seven testing locations (Figure 3) . Their widths spanned the range of [0.07, 4.76] on the logit scale, e.g., Category 2 at Adelphia, NJ, only spanned 0.59 logits, while category 8 at Stillwater, OK, was 3.54 logits. Category thresholds are generally required to be in ascending order concordant with the category numbers, i.e., ordered thresholds (Andrich, 2011). Disordered thresholds imply a higher rating may not be assigned as a turf cultivar advances along the scale. Such inconsistency of raters is usually the result of too many options or/and poor category definitions in scale development. Estimated category thresholds from all testing locations, ranging from -6.64 to 6.05, were in order. Large variations were observed in the range of category thresholds. Category thresholds at East Lansing, MI, and Stillwater, OK, spread more than 10 logits, while those in Adelphia, NJ, only spanned 4.5 logits.
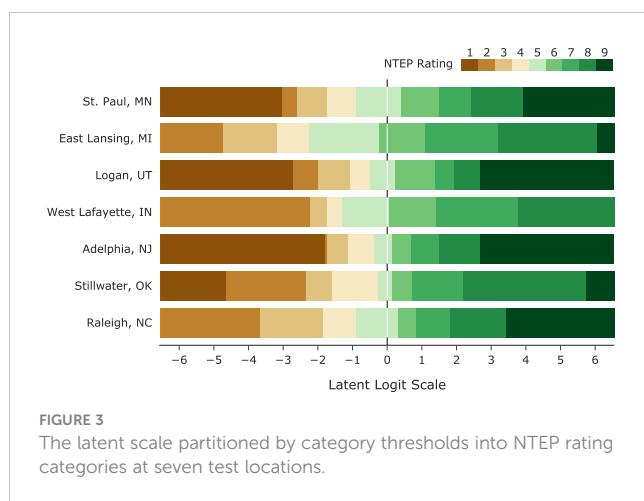
3.2.2 Rating severity

Defining category thresholds is not sufficient to answer the question of rater variation. On the constructed latent scale, category thresholds can still slide left (indicating a lenient rating event) or right (indicating a severe rating event). In many fields, severity can be treated as a constant for a given rater. That is to say, whenever the rater conducts a rating, he/she is always the same in terms of severity. However, this might not be true during the evaluation of turfgrass. For new raters, it takes time to achieve consistency; for trained raters, some may adjust their severity to credit cultivars that perform well under harsh environmental conditions or at different times of the year (personal communications with NTEP raters). Historically, there have been two sets of rating criteria for reference standards in NTEP. One is based on an optimal growth environment (e.g., light, temperature, soil moisture) and management regime (e.g., mowing height, fertilization rate), while the other is based on the actual environment or management regime. Using either criterion, the rater must idealize his/her reference standards to compare against all treatments and assign a quality score using a scale of 1 to 9. With the first criterion, we expect consistency of raters regardless of the rating time of the year since the best plot is defined considering all possible growth environments and management regimes. With the second, raters could be either severe or lenient depending on the environment or management regimes at the rating time. We examined the consistency in rating severity estimates of 10 raters who have performed more than 3 ratings across different months. For each rater, we fit a trend line for their rating severity across different months of the year using the weighted scatterplot smoothing (LOWESS) method. No strong trends were observed for raters in St. Paul, MN, West Lafayette, IN, and Adelphia, NJ, while strong seasonal patterns were seen for raters in the other four locations (Figure 4). One potential confounding factor in the current definition of rating severity is the seasonality of turfgrass quality. It is also worth noting that while the model focuses on point estimates for the average turf quality, the actual turf quality of cool-season turfgrass is not a constant; instead, it varies over time with strong annual seasonality. Unfortunately, the current data do not provide sufficient information, e.g., the exact rating dates, for investigation on how rating severity changes in response to the seasonality of turf quality. Standard deviations of rating severity per rater ranged from 0.13 to 0.97 on the logit scale. Considering the category widths, such variation in severity for a given rater could lead to changes in rating categories.

3.2.3 Field spatial variation

We implemented a Gaussian process to estimate the spatial variation within a specific trial. The traditional cultivar comparison method based on ANOVA or LLM assumes uniform growth conditions within a trial, which is hardly achievable due to heterogeneity in soil texture, seeding depth, elevation gradient, etc. Thus, removing field spatial effect is important for reliable cultivar comparison results. Figure 5 visualizes the spatial variation estimated by our model at seven testing locations, in which every pixel represents a plot as defined by row and column number. The


FIGURE 3
The latent scale partitioned by category thresholds into NTEP rating categories at seven test locations.
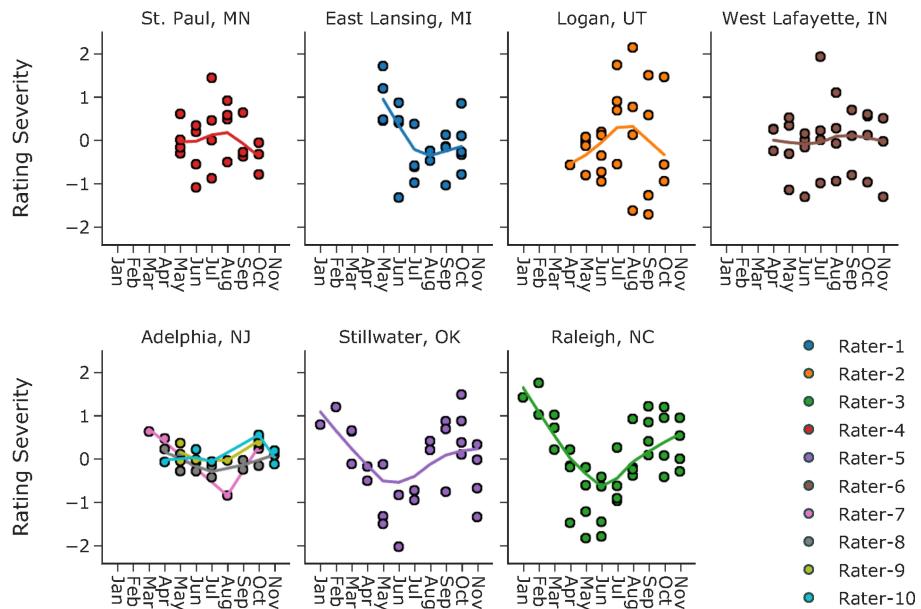
FIGURE 4

Rating severity estimates and monthly trend lines of ten raters at seven test locations.

level of spatial heterogeneity varied from trial to trial; some were higher, e.g., the trial at East Lansing, MI, while some were lower, e.g., the trial at Adelphia, NJ. Noticeably, we observed large edge effects from the trial at Logan, UT, the diagonal division from the trial at St. Paul, MN, and the localized hot spots from trials at East Lansing, MI, and Raleigh, NC. The estimated field spatial variation provided turfgrass researchers with a high-level summary of their trials, which can help improve experimental design and allow better differentiation of cultivars.

3.2.4 Cultivars comparison across testing locations

Our model quantifies and removes confounding factors at each location, i.e., rating severity and field spatial effect, allowing a more reliable and accurate cultivar comparison. An additional assumption is required for scale alignments to compare cultivars across different testing locations. We assume the average levels for a turfgrass cultivar, as perceived by raters at different NTEP testing locations, are roughly the same. In Figure 6, we compared the
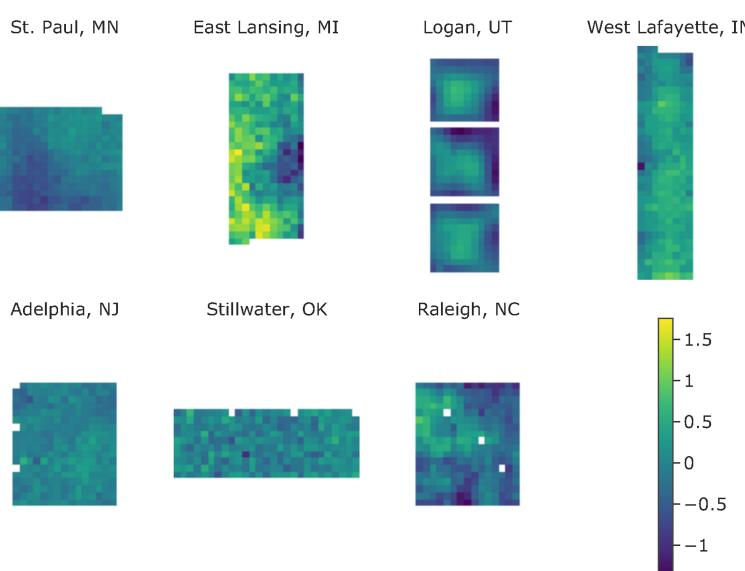

FIGURE 5

Field spatial variation at seven test locations.
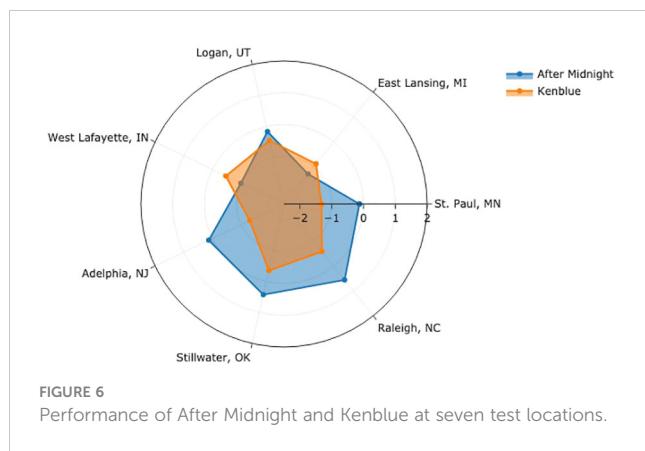
performance of two example cultivars by aligning the average levels at seven testing locations. Each angular axis represents the latent logit scale at corresponding testing locations, where zero indicates the average level. For 'After Midnight,' it performed above average at Adelphia, NJ, Stillwater, OK, and Raleigh, NC, and below average at St. Paul, MN, East Lansing, MI, Logan, UT, and West Lafayette, IN. 'Kenblue' performed below average at all locations. When comparing the two, the distance between the logit values estimates how much one cultivar is better than the other at each location. After Midnight outperformed Kenblue at all testing locations except East Lansing, MI, and West Lafayette, IN. The comparison of all evaluated cultivars can be found in Supplementary Materials and the GitHub repository.

3.2.5 Effect sizes

Effect size quantifies the strengths of relationships between variables and determines their practical importance in the study. One way to determine the effect size is by examining the percentage of variance the effects explain. Figure 7 illustrates the variance percentage explained by the model's estimated parameters. At all locations except Logan, UT, the effect of field spatial variation is the smallest of the three. In contrast, the effect of rating severity is the largest at all locations but at Adelphia, NJ. Notably, there are seven raters at Adelphia, NJ, compared with 1 to 3 raters at other locations, highlighting the importance of gathering opinions from more raters during cultivar evaluation. The percentage of variance

explained by cultivar effect varied drastically, from a merely 4% at Logan, UT, to as much as 79% at Adelphia, NJ. Quantifying and removing these confounding factors is thus essential when evaluating and comparing cultivars in field trials.

3.3 Comparison with LMM

The advantages of NTEP RSM over the currently-adopted LMM are three-folded. First, it allows the estimation of additional parameters, namely category thresholds, rating severity, and field spatial variation. All three parameters are essential for rater training, better utilization of the whole scale, and understanding of the field conditions. Second, NTEP RSM separates mean estimations of the evaluated cultivars better. To name a few of the numerous examples, Blue Gem (NAI-13-9), MVS-130, Heartland (NAI-14-187), AKB3241, and RAD 553 all received the same mean estimation of -0.261 at East Lansing, MI, from LLM, while the mean estimates from NTEP RSM were 0.030, -0.020, -0.145, -0.268, -0.580 respectively. Similar patterns were observed for DLFPS-340/3556, Paloma (PST-K13-139), DLFPS-340/3552, J-1138 at St. Paul, MN; DLFPS-340/3556, A16-2, NuRush (J-3510) at West Lafayette, IN; and DLFPS-340/3548, A16-17, Barvette HGT®, NK-1 at Logan, UT. Detailed comparison for all cultivars can be found in Among the seven test locations, the largest discrepancies between the two models' output were seen at Logan, UT. At the same time, the smallest were observed at Stillwater, OK (Table 2). It is important to highlight the robustness of the current LMM approach despite all the merits of NTEP RSM. Last but not least, RSM provides more realistic standard deviation estimations, while the currently-adopted LMM generates the same standard deviations for all cultivars at each location. Given the different genetic backgrounds of cultivars, they are unlikely to have the same standard deviations.

3.4 Parameter recovery with NTEP RSM

The highest value for $\hat{R}$ was 1.0 for all parameters and the log posterior, suggesting that all four chains have converged. As shown in Figure 8, all except three of the 95% credit intervals include zero, indicating the model's ability to recover the original values of the parameters.

3.5 Discussions

Despite the promising results, there are at least two major challenges that lie ahead for the successful implementation of the proposed model. The first and foremost is the lack of data. While NTEP has done a remarkable job of gathering, cleaning, organizing, and storing historical data on cultivar evaluation, a significant amount of valuable data are left out in this process. This includes but is not limited to rater identification, trial layout, rating dates, field gradient, etc. Luckily, researchers generally record and preserve such information at each trial location. Additional work is required to incorporate such data into the current NTEP database. Second, there are too few raters at some trial locations.
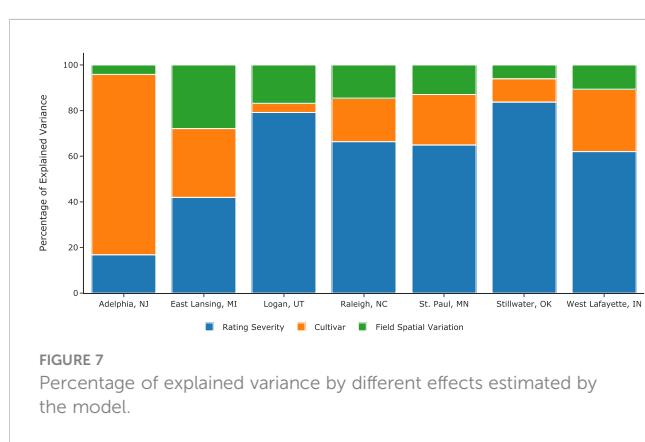
TABLE 2 Correlation coefficients between cultivar mean estimates from LMM and NTEP RSM.

Location	Correlation coefficient between LMM and NTEP RSM	
	Pearson's	Spearman's rank
St. Paul, MN	0.973614	0.970781
East Lansing, MI	0.928411	0.929173
Logan, UT	0.800883	0.756775
West Lafayette, IN	0.969092	0.955572
Adelphia, NJ	0.997716	0.997600
Stillwater, OK	0.999583	0.999022
Raleigh, NC	0.944150	0.951401

The fundamental debiasing mechanism of the proposed model is to aggregate individuals' opinions on the same cultivar into an objective and collective opinion. Multiple raters are required to ensure accurate estimations of the collective opinion on the tested cultivar. As mentioned above, one limitation of the proposed model is the absence of a seasonality component. As a cool-season turfgrass, Kentucky bluegrass thrives during the fall and early spring and slows significantly in growth during the hot summer months. The proposed model focuses on estimating the overall quality for a given cultivar over the entire testing period but cannot provide a quality estimation at a given time of the year. We tested year and month effects as independent Gaussian variables; however, as pointed out by one reviewer, it was unrealistic that months have the same effect across different years. We agree with the reviewer and are exploring better ways to improve the proposed model. A potential approach is the multiple-output Gaussian process model (Li et al., 2021) that incorporates the seasonal grown pattern of Kentucky bluegrass as a prior distribution. This requires additional information on the rating dates. Once implemented, it will allow the analysis of the temporal variation of cultivars, which caters to needs such as mixing/blending cultivars based on spring green up, comparison of cultivars on growth potential at a given time of the

year (Woods, 2013). Now that the model assumes raters are consistent in all rating event, we encourage small trial sizes at each testing location. Smaller trials reduce the risk of rater fatigue during rating, thus helping raters to maintain better consistency. For trials with too many cultivars, we recommend ratings be conducted on each replication on separate occasions instead of finishing all the plots at once. Regarding the rating scale, researchers should attempt to achieve a uniform distribution (Bond and Fox, 2013) of category thresholds. NTEP is currently working towards a data ingestion, analysis, and visualization pipeline, with the objectives to provide timely feedback to raters during the reason, to help raters to utilize the rating scale better, and to service a larger audience. NTEP also need to set standards for cultivar average, representing the zero point on the scale, such that results of cultivar comparisons across time and location are accurate and reliable.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/QhenryQ/ntep-rsm/tree/main/model_data.

Author contributions

YQ conceived the idea, developed the model, performed the analysis, and took the lead in writing the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors are grateful for the generous help from Dr. Cale A. Bigelow of Purdue University, Dr. Stacy A. Bonos of Rutgers University, Dr. Leah Brilman of DLF Pickseed, Dr. Andrea Payne Connally of Oklahoma State University, Ms. Christine Knisley of National Turfgrass Evaluation Program, Dr. Kevin W. Frank of Michigan State University, Mr. Paul Harris of Utah State University,
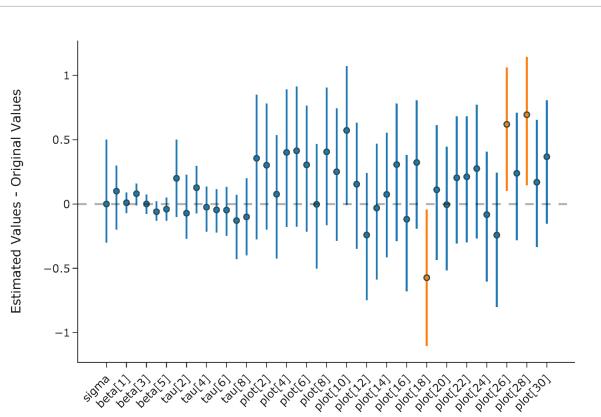
FIGURE 8
Mean estimation and 95% credit interval for the difference between estimated values and original values of the parameters.

Mr. Andrew Hollman of University of Minnesota - Twin Cities, Dr. Paul G. Johnson of Utah State University, Dr. Dennis L. Martin of Oklahoma State University, Dr. Grady L. Miller of North Carolina State University, Dr. Phillip L. Vines of the University of Georgia. We also want to express our gratitude to two reviewers whose comments helped improve and clarify this manuscript.

Conflict of interest

YQ and KM are both employed by NTEP.

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1135918/full#supplementary-material

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* 43, 561–573. doi: 10.1007/BF02293814

Andrich, D. (2011). Rating scales and rasch measurement. *Expert Rev. Pharmacoecon. Outcomes Res.* 11, 571–585. doi: 10.1586/erp.11.59

Andrich, D., and Luo, G. (2003). Conditional pairwise estimation in the rasch model for ordered response categories using principal components. *J. Appl. Meas.* 4, 205–221.

Bond, T. G., and Fox, C. M. (2013). *Applying the rasch model: fundamental measurement in the human sciences* (Psychology Press).

Bürkner, P.-C., and Vuorre, M. (2019). Ordinal regression models in psychology: a tutorial. *Adv. Methods Pract. psychol. Sci.* 2, 77–101. doi: 10.1177/2515245918823199

Ebdon, J., and Gauch, H.Jr. (2002a). Additive main effect and multiplicative interaction analysis of national turfgrass performance trials: i. interpretation of genotype× environment interaction. *Crop Sci.* 42, 489–496. doi: 10.2135/cropsci2002.4890

Ebdon, J., and Gauch, H.Jr. (2002b). Additive main effect and multiplicative interaction analysis of national turfgrass performance trials: II. Cultivar recommendations. *Crop Sci.* 42, 497–506. doi: 10.2135/cropsci2002.4970

Hollingworth, H. L. (1910). The central tendency of judgment. *J. Philosophy Psychol. Sci. Methods* 7, 461–469. doi: 10.2307/2012819

Jones, L. V., Peryam, D. R., and Thurstone, L. L. (1955). Development of a scale for measuring soldiers' food preferences. *Food Res.* 20, 512–520. doi: 10.1111/j.1365-2621.1955.tb16862.x

Jones, L. V., and Thurstone, L. L. (1955). The psychophysics of semantics: an experimental investigation. *J. Appl. Psychol.* 39, 31. doi: 10.1037/h0042184

Lee, D., Carpenter, B., Li, P., Morris, M., Betancourt, M., Maverick, G., et al. (2017). *Stan-dev/stan: v2.17.1.* doi: 10.5281/zenodo.1101116

Li, D., Jones, A., Banerjee, S., and Engelhardt, B. E. (2021). Multi-group gaussian processes. doi: 10.48550/arXiv.2110.08411

Liddell, T. M., and Kruschke, J. K. (2018). Analyzing ordinal data with metric models: what could possibly go wrong? *J. Exp. Soc. Psychol.* 79, 328–348. doi: 10.1016/j.jesp.2018.08.009

Lim, J., Wood, A., and Green, B. G. (2009). Derivation and evaluation of a labeled hedonic scale. *Chem. Senses* 34, 739–751. doi: 10.1016/j.foodqual.2011.05.008

Moskowitz, H. R. (1971). The sweetness and pleasantness of sugars. *Am. J. Psychol.* 84, 387–405. doi: 10.2307/1420470

Moskowitz, H. R. (1977). Magnitude estimation: notes on what, how, when, and why to use it. *J. Food Qual.* 1, 195–227. doi: 10.1111/j.1745-4557.1977.tb00942.x

Moskowitz, H. R. (1980). Psychometric evaluation of food preferences. *Foodservice Res. Int.* 1, 149–167. doi: 10.1111/j.1745-4506.1980.tb00252.x

Moskowitz, H. R., and Sidel, J. L. (1971). Magnitude and hedonic scales of food acceptability. *J. Food Sci.* 36, 677–680. doi: 10.1111/j.1365-2621.1971.tb15160.x

Parducci, A., and Wedell, D. H. (1986). The category effect with rating scales: number of categories, number of stimuli, and method of presentation. *J. Exp. Psychol.: Hum. Percept. Perform.* 12, 496. doi: 10.1037/0096-1523.12.4.496

Peryam, D. R., and Girardot, N. F. (1952). Advanced taste-test method. *Food Eng.* 24, 58–61.

Peryam, D. R., and Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technol.* 11, 9–14. doi: 10.1007/BF02638783

Seabold, S., and Perktold, J. (2010). "Statsmodels: econometric and statistical modeling with python," in *9th Python in Science Conference* (Austin, Texas: SciPy). doi: 10.25080/majora-92bf1922-011

Seddon, M., Marshall, M., Campbell, S., and Roland, M. (2001). Systematic review of studies of quality of clinical care in general practice in the uk, australia and new zealand. *BMJ Qual. Saf.* 10, 152–158. doi: 10.1126/science.103.2684.677

Stevens, S. S., and Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *J. Exp. Psychol.* 54, 377. doi: 10.1037/h0043680

Woods, M. (2013). *Using temperature to predict turfgrass growth potential (gp) and to estimate turfgrass nitrogen use* (Bangkok: Asian Turfgrass Publication.[Google Scholar]).

Xie, Y., Farhadloo, M., Guo, N., Shekhar, S., Watkins, E., Kne, L., et al. (2022). Ntep-db 1.0: a relational database for the national turfgrass evaluation program. *Int. Turfgrass Soc. Res. J.* 14, 316–332. doi: 10.1002/its2.76

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

Frontiers in
Plant Science

frontiers | Research Topics