2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Algorithm Research for Crop Pest and Disease Identification Based on Improved YOLOv8

YueHong Lin[1],Chen Dong[2]✉, ShuTingWei[2]

[1]School of Huaxin Software , Tianjin University of Technology, Tianjin,China.
[2]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin,China.
dongc@tjut.edu.cn

Abstract. To address the issues of insufficient feature extraction in complex environments, missed detections of small-scale pests and diseases, and inadequate multi-scale feature fusion in crop pest and disease detection, this paper proposes an improved YOLOv8-based pest and disease identification algorithm.First, to enhance feature extraction in complex scenarios, the Swin Transformer module was introduced into the YOLOv8 backbone network. Leveraging its hierarchical structure and Shifted Window Multi-Head Self-Attention, the model's ability to capture global pest and disease features was strengthened. Second, to mitigate missed detections of small-scale pests and diseases, an SE attention module was added to the Neck, enabling adaptive channel-wise feature weighting to enhance feature representation. Finally, the YOLOv8 Concat module was replaced with BiFPN, which uses a learnable bi-directional fusion strategy to optimize cross-scale feature interactions. Experimental results showed that the improved YOLOv8 model excelled in detecting 23 crop pests and diseases, achieving 96.6% precision and 98.2% mAP50. It also maintained high accuracy and efficiency under challenging conditions like overcast or strong lighting, demonstrating strong application potential.

Keywords. Pest and disease detection, YOLOv8, Swin Transformer, BiFPN, Multi-scale feature fusion

✉ Corresponding author

1.Introduction

In modern agricultural production, pests and diseases pose a significant threat to crop yield and quality. Traditional manual detection methods are costly, inefficient, and highly subjective, making them unsuitable for large-scale applications. High-precision techniques such as spectral imaging [1] are limited by their high costs. Deep learning-based object detection algorithms offer new solutions for intelligent pest and disease detection but still face challenges, including insufficient feature extraction in complex environments, missed detections of small-scale pests, and inadequate multi-scale feature fusion.Deep learning detection methods are generally categorized into two-stage and single-stage algorithms [2]. Two-stage methods achieve higher accuracy but are slower, whereas single-stage methods offer slightly lower accuracy but faster detection speed, making single-stage methods more suitable for agricultural scenarios with limited computing resources and real-time requirements. This study adopts the YOLOv8 single-stage detection algorithm and proposes several improvements. The main contributions are as follows:

1. To address the challenge of insufficient feature extraction caused by complex lighting variations, weed interference, and target occlusion in pest and disease detection, the Swin Transformer module is introduced into the YOLOv8 backbone network. This enhances the model's global feature extraction capability and improves its ability to distinguish pests and diseases from complex backgrounds.
2. To mitigate the issue of missing small-scale pests and diseases, the SE attention module is incorporated into the Neck part of the model. This module adaptively enhances key channel features, thereby improving the detection accuracy of small targets.
3. To improve multi-scale feature fusion, which is often insufficient due to variations in pest and disease size across different growth stages and viewing angles, the Concat module in YOLOv8 is replaced with the BiFPN module. This enables bidirectional weighted feature fusion, enhancing multi-scale detection performance and making the model more adaptable to the diversity of pests and diseases.

2.Related Works

2.1.Traditional Machine Learning Methods

Traditional machine learning methods primarily rely on handcrafted feature extraction combined with classifiers for pest and disease detection. These studies typically design features based on color, texture, and shape, utilizing classification algorithms such as SVM, Decision Trees, or KNN for identification [3].

Pusadan et al. [4] employed the HSV color feature extraction method and the KNN algorithm to detect pests and diseases in cocoa pods. When using the hold-out method, the best accuracy achieved was 84.44% at k=5. However, this approach suffered from issues such as sensitivity to image quality, imbalanced feature value distributions, and variations in object input positions.Kale and Shitole [5] integrated KNN, SVM, and Random Forest for crop pest and disease detection. While KNN demonstrated good stability in handling complex data, its overall accuracy was lower than that of SVM, with weaker generalization capability.Chanda and Biswas [6] utilized K-means clustering and the Particle Swarm Optimization (PSO) algorithm in combination with a Backpropagation Neural Network (BPNN) for disease detection. Although their method achieved an accuracy of 96.2% in certain pest and disease identification tasks, it incurred high computational costs and struggled with high-dimensional data processing.

Traditional machine learning is limited by manually designed features, struggling with complex pest morphology, background interference, and multi-scale targets. Its detection process is inefficient and unsuitable for large-scale agriculture. In contrast, deep learning, with end-to-end learning and automatic feature extraction, overcomes these limitations.

2.2.Deep Learning Methods

In recent years, deep learning has emerged as a key technology for crop pest and disease detection due to its ability to perform automatic feature extraction and high-dimensional modeling. Convolutional Neural Networks (CNNs) and their variants, such as ResNet [7], MobileNet [8], and EfficientNet [9], as well as object detection algorithms like Faster R-CNN [10] and the YOLO series , have been widely applied in this domain. However, these methods still present certain limitations in practical applications.

Sood and Singh [11] employed convolutional neural network models based on ResNet50 and VGG16 for wheat rust detection. Experimental results indicated that VGG16 achieved high detection accuracy when provided with a sufficient number of samples. However, its adaptability to variations in lighting conditions and complex field backgrounds was poor, making it susceptible to noise interference and limiting its effectiveness in detecting small-scale pests and diseases.Priyadharshini et al. [12] utilized Faster R-CNN combined with VGG16 to detect tomato leaf pests and diseases, achieving an accuracy of 98% with relatively fast detection speeds, making it suitable for certain pest and disease scenarios. However, as a two-stage detection algorithm, this approach has a high model complexity and significant computational resource consumption, making it difficult to meet real-time detection requirements. Furthermore, it struggles with recognizing pests and diseases across multiple scales.Yang et al. [13] designed a lightweight YOLOv5-based model for detecting bayberry fruit pests and diseases, improving inference speed and expanding the receptive field. However, the model primarily focused on fruit detection and did not account for early-stage pest and disease characteristics on leaves, limiting its effectiveness in early warning applications.

In summary, although deep learning has made significant progress in crop pest and disease detection, further improvements are needed to enhance adaptability to complex environments, improve small-scale target detection accuracy, and optimize multi-scale feature fusion. These advancements are essential to meet the demands for high precision, real-time performance, and broad applicability.

3. Disease And Insect Detection Overview

3.1. Datasets

This study utilized a publicly available dataset from the Roboflow platform, which comprised 23,123 images covering 23 different types of crop pests and diseases. The dataset was randomly divided into training, validation, and test sets in a ratio of 15:1:1. The specific classification and the number of images for each pest and disease category are illustrated in Figure 1.

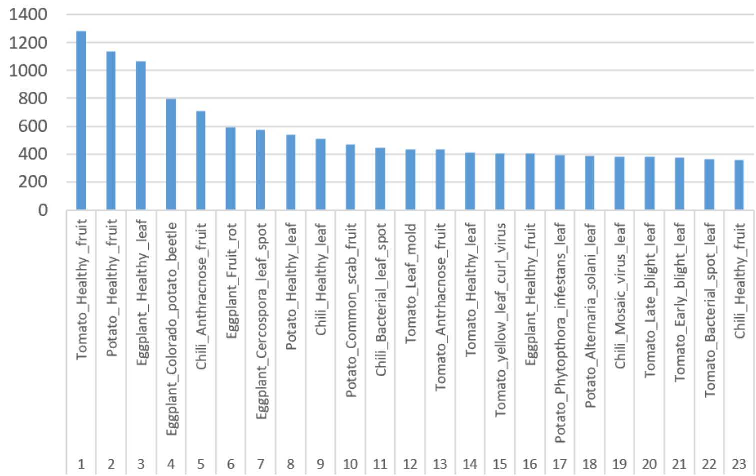


Fig. 1. Classification of pest and disease datasets and their quantities

In the pest and disease identification task, data preprocessing is performed to enhance the model's generalization ability in complex environments, such as rainy conditions, overexposed settings, and low-light scenarios. This preprocessing improves the model's robustness and recognition accuracy. Figure 2 presents a selection of images from the dataset.

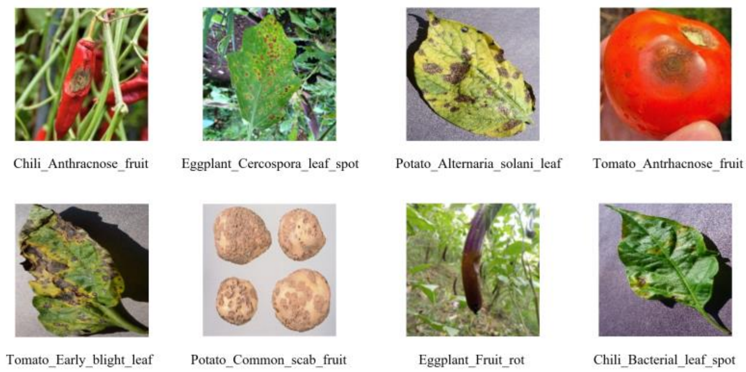


Fig. 2. Display of partial crop pest and disease datasets

During preprocessing, images are aligned, cropped (10–85% horizontal, 4–82% vertical), and resized to 640×640 to reduce background interference. The HVI-CIDNet[14] model enhances visibility in low light, improving feature contrast. To boost robustness, each sample undergoes two augmentations, including 90° rotation, 20% grayscale, brightness adjustment (-15% to +15%), and bounding box rotation (-10° to +10°), aiding recognition under varying conditions.Figure 3 illustrates the effects of data augmentation on crop disease images.

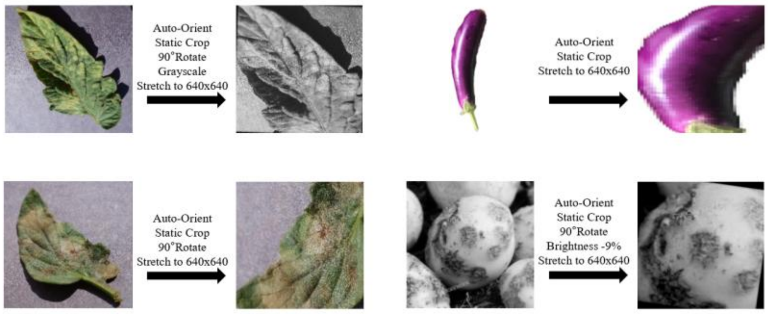


Fig. 3. Examples of image data augmentation for crop diseases

3.2. YOLOv8 Introduction

YOLOv8 [15] is an optimized single-stage object detection algorithm that enhances accuracy and flexibility. Its architecture includes three components: Backbone, Neck, and Head. The Backbone uses a lightweight C2f module for better feature extraction, while the Neck employs an improved Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) for multi-scale fusion. The decoupled Head separates classification and regression tasks, improving prediction accuracy. YOLOv8 also adopts an anchor-free design, simplifying training and reducing hyperparameter dependency. The loss function integrates Task-Aligned Assigner and Distribution Focal Loss for better target matching, and the Spatial Pyramid Pooling-Fast (SPPF) module boosts adaptability to objects of various sizes. Overall, YOLOv8 balances accuracy, flexibility, and speed, making it ideal for real-time pest and disease detection.

3.3. Improved YOLOv8

To address the limitations of pest and disease feature extraction in complex environments, including insufficient detection of small-scale pests and inadequate multi-scale feature fusion, this study optimized the YOLOv8 model as follows:

First, the Swin Transformer was integrated into the backbone to enhance feature extraction by combining convolution with window-based self-attention. It alternates between Window Multi-Head Self-Attention (W-MSA) and Shifted Window Multi-Head Self-Attention (SW-MSA), capturing global and local features while reducing complexity. This improved the model’s ability to identify pest-affected areas in challenging conditions like uneven lighting or background similarity, enhancing pest and disease feature extraction.Second, the SE attention module was incorporated into the Neck structure to improve the detection of small-scale pests and edge regions. By introducing the SE attention module, the model leverages global pooling to capture channel-wise dependencies, adaptively adjusting feature channel weights to emphasize critical pest-related features while suppressing background noise. This enhancement strengthened the detection of weakly represented pests, improving the model’s performance in early-stage pest and disease identification tasks.Finally, BiFPN replaced Concat for multi-scale feature fusion to address pest scale variations. YOLOv8’s simple concatenation struggles to integrate hierarchical features, reducing detection accuracy. BiFPN introduced a learnable weighted fusion strategy, enhancing small pest detection and improving large pest recognition. Its bidirectional interaction between top-down and bottom-up pathways ensured more effective feature integration across different scales.The improved YOLOv8 network architecture is illustrated in Figure 4.

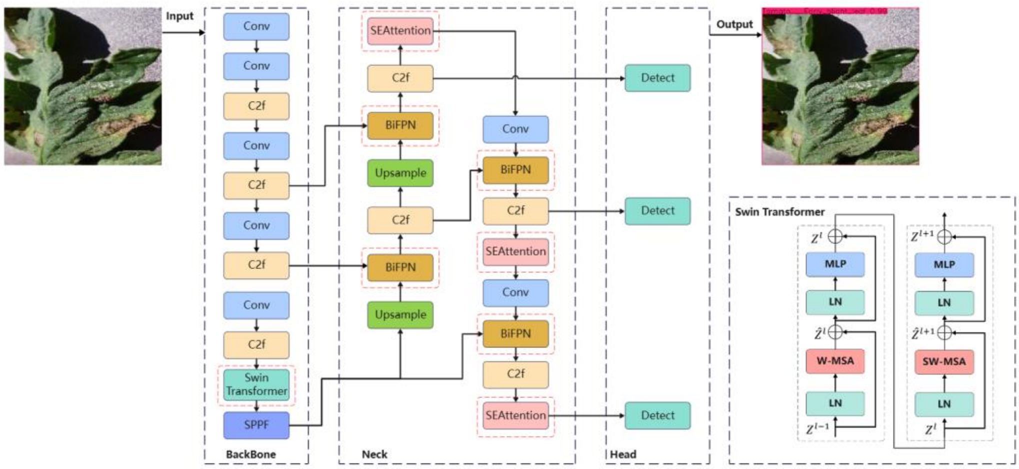


Fig. 4. Improved YOLOv8 network architecture

3.3.1. Swin-Transform

In natural farmland environments, illumination changes, weed interference, and occlusion hinder feature extraction, reducing detection accuracy. Traditional CNNs struggle with large-scale variations and uneven lighting due to limited receptive fields. To overcome this, Swin Transformer [16] was integrated into the YOLOv8 Backbone, using W-MSA and SW-MSA to expand the receptive field. This enabled the model to capture both local and global features, improving the recognition of pest and disease contours in complex scenarios.

In Swin Transformer, pest and disease images are divided into non-overlapping windows, and features within each window are processed using W-MSA, reducing computational complexity and enabling linear scaling with image size. However, this window partitioning limits interaction between features from different windows. To overcome this, Swin Transformer introduces SW-MSA, which shifts window positions between layers to connect previously isolated windows. This expanded the receptive field and improves the model's ability to capture global patterns, enhancing recognition of pest and disease features. The process is shown in Figure 5.

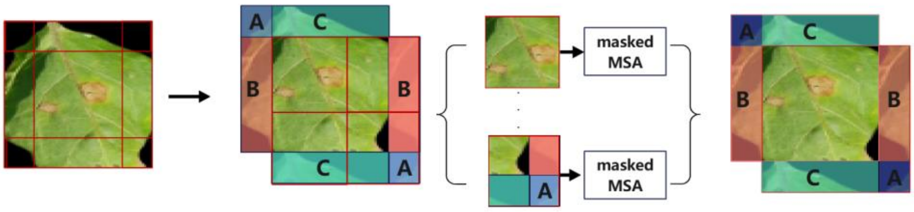


Fig. 5. Shifting window attention mechanism on pest and disease leaf images

Within each window, self-attention weights are computed using the scaled dot-product attention mechanism as follows:

$$Q = X \cdot W_Q \quad (1)$$

$$K = X \cdot W_K \quad (2)$$

$$V = X \cdot W_V \quad (3)$$

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_K}} + B\right)V \quad (4)$$

Where X represents the input feature matrix within the window; W_Q、W_K and W_V are the linear transformation matrices for the query, key, and value, respectively; d_k denotes the dimension of the key vector, and $\sqrt{d_k}$ is used for scaling to prevent excessively large gradients; B is the relative position encoding within the window, which facilitates modeling of local relationships.

The Swin Transformer enhanced multi-scale feature extraction, making the model more adaptable to complex environments with varying pest and disease sizes, shapes, and structures. This improved its ability to distinguish pest-affected regions from the background and extract pest-related features in intricate settings.

3.3.2. SE-Attention Module

In crop pest and disease detection, complex backgrounds and subtle, blurred pest features make traditional convolution-based methods prone to noise interference, leading to missed detections or false positives. To address this, the Squeeze-and-Excitation (SE) attention module [17] was integrated into YOLOv8's Neck. The SE module uses Global Average Pooling (GAP) to extract global channel information and adaptively adjusts channel weights based on pest and disease texture and color. This enhanced key feature response, such as disease edges, while suppressing background noise, improving detection performance in complex agricultural environments.

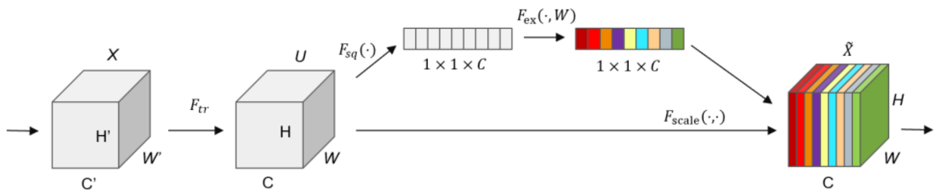


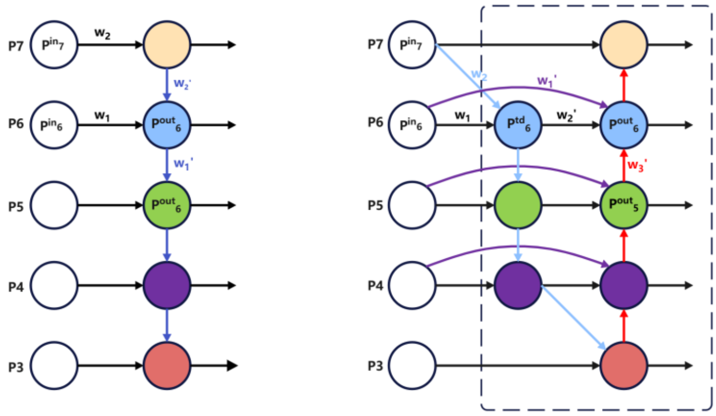
Fig. 6. SE Attention Module

The SE attention module consists of three components: the Squeeze stage, the Excitation stage, and the Scale stage, as illustrated in Figure 6.

In the Squeeze stage, GAP captures global features while reducing redundancy, aiding disease texture detection. The Excitation stage assigns channel importance using ReLU and Sigmoid, enhancing pest and disease-related features while suppressing background noise. In the Scale stage, attention weights refine feature representation, improving disease detection in complex environments. For instance, the SE module highlights abnormalities in color or texture within diseased areas, which enhanced the model’s accuracy. The SE attention module significantly improved the model’s ability to detect small-scale pest and disease symptoms, identify subtle color differences, and maintain robustness in noisy backgrounds, providing a strong foundation for accurate detection in complex agricultural settings.

3.3.3. BiFPN

In crop pest and disease detection, accurate identification of both small and large affected areas is challenging due to significant scale differences. Traditional feature fusion methods, like YOLOv8's Concat module, use fixed fusion weights, which limit performance for small or large targets. To address this, the Concat module was replaced with the Bidirectional Feature Pyramid Network (BiFPN) [18]. BiFPN improves upon the traditional Feature Pyramid Network (FPN) by incorporating learnable weights, allowing dynamic adjustment of feature map importance based on target size. This enables better fine-grained detection for small-scale regions and improved boundary recognition for large-scale areas. The BiFPN structure is shown in Figure 7.



(a) FPN (b) BiFPN

Fig. 7. Comparison of FPN and BiFPN architectures

BiFPN enhances multi-scale perception in complex environments, handling overlapping leaves, uneven lighting, and diverse pest and disease morphologies while balancing accuracy and efficiency. Replacing the original module with BiFPN significantly improved detection of small pest regions and irregular disease boundaries.

4.Experiments

4.1. Equipment And Parameter Settings

The experimental operating system in this study was Ubuntu 20.04, with PyTorch as the deep learning framework for model development. The specific training environment is detailed in Table 1.

Table 1. Experimental environment configuration

Parameters	Configuration
CPU	12 vCPU Intel(R) Xeon(R) Silver 4214R CPU
GPU	RTX 3080 Ti
GPU memory size	12GB
Operating systems	ubuntu20.04
Deep learning architecture	PyTorch 1.11.0 + Cuda 11.3 +

The training parameters were set as follows: the input image size was 640×640, the batch size was 16, and the initial learning rate was 0.01. The model was trained for a total of 120 epochs. The detailed parameter settings are shown in Table 2.

Table 2. Training parameter configurations

Parameters	Value
Task	Detect
Epochs	120
Batch Size	16
Image Size	640
Optimizer	SGD

4.2. Comparative Experiments

To validate the effectiveness of the proposed model, Faster R-CNN, SSD, YOLOv5, YOLOv8, and the improved YOLOv8 model were selected for comparison experiments under the same training environment using the Roboflow crop pest and disease dataset. To evaluate the detection performance of the improved YOLOv8 model, relevant metrics such as Precision, Recall, and mAP were used as evaluation indicators. The training process is illustrated in Figure 8.

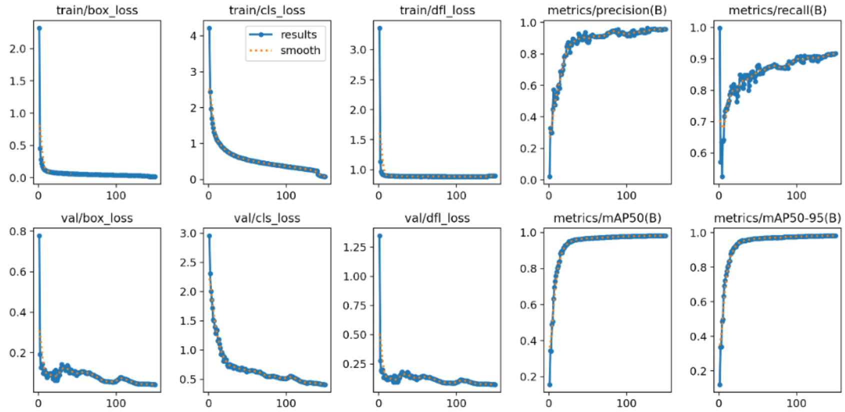


Fig. 8. Training process of improved YOLOv8 model

The experimental results indicated that although Faster R-CNN achieved a high recall rate (87.65%) and could detect most pest and disease targets, it suffered from low precision (56.89%), leading to frequent misclassification of background regions as pest or disease areas. Additionally, its inference speed was slow (23.94 FPS), making it unsuitable for real-time agricultural applications. The SSD model demonstrated a much faster detection speed (119.72 FPS), making it suitable for preliminary screening in rapid detection scenarios. However, it had a relatively low recall rate (67.83%), resulting in significant missed detections, particularly in complex backgrounds and small pest or disease regions. This limitation made it challenging to meet the requirements of precise pest and disease control. The detailed experimental results are shown in Table 3.

Table 3. Comparison of pest and disease detection performance across different models

Network	Precision/%	Recall/%	mAP50/%	FPS
Faster R-CNN	56.89	87.65	84.3	23.94
SSD	84.37	67.83	79.4	119.72
YOLOv5	95.6	92.0	98.4	333.87
YOLOv8	94.8	89.8	97.8	275.78
Improved YOLOv8	96.6	91.8	98.2	250.48

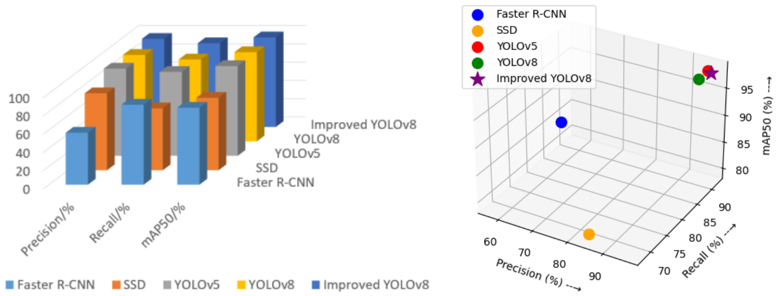
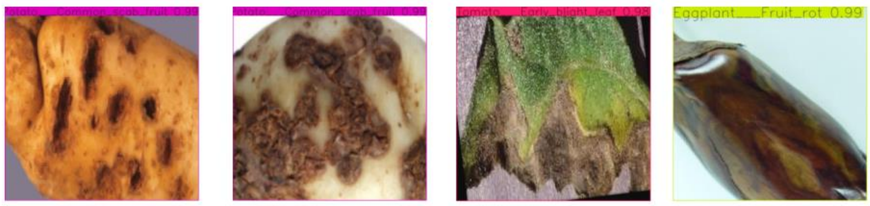


Fig. 9. Comparison of parameter bar chart and scatter plot for different network models

YOLOv5 achieved balanced performance with high precision (95.6%), recall (92.0%), and mAP@50 (98.4%) but struggled under overexposure and occlusion. The standard YOLOv8 improved both speed (275.78 FPS) and precision (94.8%), offering certain advantages in multi-scale pest and disease detection. However, it still encountered missed detections when dealing with small-scale pest and disease regions and complex textured backgrounds. The improved YOLOv8 achieved high precision (96.6%), mAP@50 (98.2%), and recall (91.8%). The Swin Transformer enhanced feature extraction in complex backgrounds, while the SE attention module strengthened small and low-contrast pest detection, reducing missed cases. BiFPN optimized bidirectional multi-scale fusion, improving detection of both large and small pests. Additionally, the model maintained high accuracy with an inference speed of 250.48 FPS, enabling real-time detection and large-scale agricultural deployment. Some detection results of the improved YOLOv8 model are shown in Figure 10.



(a) Potato__Common_scab_fruit (b) Potato__Common_scab_fruit (c) Tomato__Early_blight_leaf (d) Eggplant__Fruit_rot

Fig. 10. Detection results of the improved YOLOv8 model

4.3. Ablation Experiment and Result

To evaluate the specific contributions of each module to the model's performance, an ablation study was conducted. The Swin Transformer, SE attention module, and

BiFPN were individually removed or retained to analyze their individual and combined effects on model performance, as shown in Table 4 and Figure 11. Furthermore, to assess the impact of each module on feature extraction, Grad-CAM was used to generate heatmaps that visualize the attention regions of different model configurations.This offers a more intuitive basis for analysis, as shown in Figure 12.

Table 4. Ablation experiment results

Network	Precision/%	Recall/%	mAP50	mAP50-95	FPS
Yolov8	94.8	89.8	97.8	97.8	275.78
Improved YOLOv8	96.6	91.8	98.2	98.2	250.48
SEA	94.1	92.6	98.2	98.2	285.51
BiFPN	95.9	93.4	98.6	98.5	242.19
Swin Transformer	94.3	93.0	98.2	98.2	318.46
SEA+BinFPN	95.7	92.8	98.4	98.4	267.71
SEA+SwinTransformer	94.7	92.5	97.8	97.8	290.87
BiFPN+SwinTransformer	95.6	92.0	98.1	98.0	277.00

After introducing the SE attention module alone, the recall rate increased to 92.6%, and mAP@50 reached 98.2%, indicating its significant effect in enhancing small-scale pest and disease feature representation. The heatmap visualization showed that the model's attention was more concentrated on the core regions of pests and diseases. However, it also exhibited over-activation in some non-pest areas, leading to a slight decrease in precision.

Fig. 11. Comparison of detection performance across different network models

After incorporating BiFPN, the model's precision increased to 95.9%, and mAP@50 reached 98.6%. The heatmap visualization showed a stronger response in boundary disease regions and small target areas, indicating that BiFPN enhanced multi-scale feature interactions. This improvement was particularly beneficial for pest and disease detection in complex backgrounds and scenarios with significant variations in target scale.

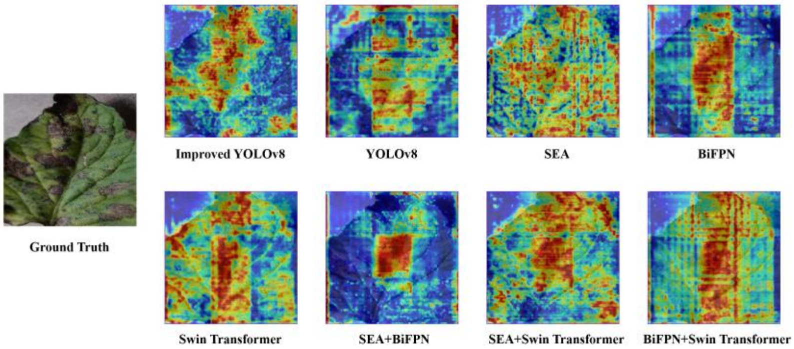


Fig. 12. Heatmap of pest and disease detection generated by the Grad-CAM

The introduction of the Swin Transformer increased the recall rate to 93.0% and achieved the fastest inference speed (318.46 FPS). It excelled in large-scale disease detection and global feature extraction but struggled with small targets.

When different modules were combined, the SE+BiFPN combination achieved a balanced performance in terms of accuracy (95.7%), mAP@50 (98.4%), and boundary detection. The SE+Swin Transformer combination enhanced the model's global recognition capability but slightly suppressed small target detection. The BiFPN+Swin Transformer combination effectively integrated multi-scale perception with global awareness.The improved YOLOv8, incorporating all three components, optimized various evaluation metrics, including accuracy (96.6%), recall (91.8%), and mAP@50 (98.2%). Heatmaps further indicated that the model could precisely focus on pest and disease regions with clear boundary delineation, making it well-suited for early detection of plant diseases and pests in complex environments, small-scale pest identification, and precise detection of targets across different scales.

5. Conclusions

To tackle challenges in crop pest and disease detection, such as feature extraction difficulties, missed small-scale cases, and inadequate multi-scale fusion, this study proposed an improved YOLOv8-based model. It integrated Swin Transformer for global feature extraction, the SE attention module for better small-scale detection, and

BiFPN for optimized multi-scale fusion. In detecting 23 pest and disease categories, the model achieved 96.6% accuracy, 91.8% recall, and 98.2% mAP@50, outperforming traditional models.

Experiments showed that the model excelled in real-time detection, even in challenging conditions like rain or overexposure, making it a reliable solution for large-scale monitoring. Future research will explore lightweight architectures, such as Depthwise Separable Convolution, to reduce computational costs, and advanced attention mechanisms, like Global-Local Adaptive Attention, to further enhance detection in complex environments.

Acknowledgement: The authors gratefully acknowledge the financial supports by the National Natural Science Foundation of China under Grant (62306212);College Student Innovation and Entrepreneurship Training Program(202410060007).

References:

1. Rayhana R, Ma Z, Liu Z, et al. A review on plant disease detection using hyperspectral imaging[J]. IEEE Transactions on AgriFood Electronics, 2023, 1(2): 108-134.
2. Mittal M, Gupta V, Aamash M, et al. Machine learning for pest detection and infestation prediction: A comprehensive review[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2024, 14(5): e1551.
3. Chithambarathanu M, Jeyakumar M K. Survey on crop pest detection using deep learning and machine learning approaches[J]. Multimedia Tools and Applications, 2023, 82(27): 42277-42310.
4. Pusadan M Y, Abdullah A I. k-Nearest neighbor and feature extraction on detection of pest and diseases of cocoa[J]. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 2022, 6(3): 471-480.
5. KALE M R S R, SHITOLE M R S. Analysis of crop disease detection with SVM, KNN and random forest classification[J]. Information Technology in Industry, 2021, 9(1): 364-372.
6. Chanda M, Biswas M. Plant disease identification and classification using back-propagation neural network with particle swarm optimization[C]//2019 3rd international conference on trends in electronics and informatics (ICOEI). IEEE, 2019: 1029-1036.

7. Kumar V, Arora H, Sisodia J. Resnet-based approach for detection and classification of plant leaf diseases[C]//2020 international conference on electronics and sustainable communication systems (ICESC). IEEE, 2020: 495-502.
8. Ashwinkumar S, Rajagopal S, Manimaran V, et al. Automated plant leaf disease detection and classification using optimal MobileNet based convolutional neural networks[J]. Materials Today: Proceedings, 2022, 51: 480-487.
9. Kunduracıoğlu İ, Paçal İ. Deep Learning-Based Disease Detection in Sugarcane Leaves: Evaluating EfficientNet Models[J]. Journal of Operations Intelligence, 2024, 2(1): 321-235.
10. Bari B S, Islam M N, Rashid M, et al. A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework[J]. PeerJ Computer Science, 2021, 7: e432.
11. Sood S, Singh H. An implementation and analysis of deep learning models for the detection of wheat rust disease[C]//2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2020: 341-347.
12. Priyadharshini G, Dolly D R J. Comparative investigations on tomato leaf disease detection and classification using CNN, R-CNN, fast R-CNN and faster R-CNN[C]//2023 9th International conference on advanced computing and communication systems (ICACCS). IEEE, 2023, 1: 1540-1545.
13. Yang C Y, Liu J, He J T. A lightweight waxberry fruit detection model based on YOLOv5[J]. IET Image Processing, 2024, 18(7): 1796-1808.
14. https://doi.org/10.48550/arXiv.2502.20272
15. Varghese R, Sambath M. Yolov8: A novel object detection algorithm with enhanced performance and robustness[C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, 2024: 1-6.
16. Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992-10002
17. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
18. M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10778-10787