Highlights

MSN-Mamba-YOLO: A Novel Detection Framework for Multi-Crop and Multi-Disease Identification

CV Radhakrishnan,Han Theh Thanh,CV Rajagopal,Rishi T.

- Research highlights item 1
- Research highlights item 2
- Research highlights item 3

MSN-Mamba-YOLO: A Novel Detection Framework for Multi-Crop and Multi-Disease Identification ⋆,⋆⋆

Sir CV Radhakrishnan[a,c,*,1] (Researcher), Han Theh Thanh[b,d], CV Rajagopal Jr[b,c,2] (Co-ordinator) and Rishi T.[a,c,**,1,3]

[a]Elsevier B.V., Radarweg 29, Amsterdam, 1043 NX, The Netherlands
[b]Sayahna Foundation, Jagathy, 695014, Trivandrum, India
[c]STM Document Engineering Pvt Ltd., Mepukada, Malayinkil, 695571, Trivandrum, India

ARTICLE INFO

Keywords:
Multi-crop and Multi-class disease detection
Mamba-YOLO
MobileMamba
YOLO11
Cross-domain detection

Abstract

Agricultural diseases pose a serious threat to crop yield and quality. Although deep learning has made significant progress in disease detection, multi-crop and multi-class disease detection still faces challenges such as feature similarity, poor environmental adaptability, and high computational costs. To address these issues, we propose a novel lightweight detection model based on Mamba-YOLO. Specifically, we construct an efficient feature extractor using MobileMamba and integrate the CSPSA module from YOLO11 to mitigate multi-class feature similarity. In the neck, we design a dual-path concurrent global scanning mechanism (S6 and serpentine scanning) to enhance environmental adaptability, while the original detection head is replaced with a lightweight YOLO11 head to improve efficiency. On the augmented FieldPlant dataset, our model achieved a mAP@50–95 of 80.5%, representing an improvement of 0.7% compared to the baseline model, while reducing the number of parameters and computational cost by 22% and 30%, respectively. Compared with YOLO11s, the mAP@50–95 was 0.1% higher, with a reduction of 51% in parameters and 60% in computational cost. Furthermore, we validate its cross-domain detection capability on tea and eggplant disease datasets. Our study significantly improves the universality and scalability of agricultural disease monitoring and provides strong support for the development of large-scale intelligent agricultural disease early-warning systems.

1. Introduction

One of the major challenges facing agriculture is plant diseases. With the latest advances in deep neural network technology, researchers have been able to significantly improve the efficiency of object recognition and detection systems (Rezk et al., 2022).However, most current studies on agricultural diseases focus on targeted improvements for a single crop, while in practice, the simultaneous infestation of multiple crop diseases poses a serious threat to global agricultural production and food security. These factors may lead to severe yield losses, reduced crop quality, and an increased reliance on chemical interventions (Bangal et al., 2025). Therefore, there is an urgent need for a method capable of detecting multiple crop diseases to address this challenge.

With the rapid development of deep learning, significant progress has been made in plant disease detection. However, its practical application still faces several challenges:(1) Feature similarity is particularly severe in multi-crop and multi-disease detection tasks (Wang et al., 2024a). Diseases on different crops may exhibit highly similar visual symptoms, such as brown spots or leaf chlorosis, which share similar color, texture, and morphological patterns, making them difficult to distinguish (Upadhyay et al., 2025). Moreover, subtle differences among multiple diseases within the same crop further increase the classification difficulty (Barbedo, 2018).(2) The accuracy of existing models is insufficient when applied to field-acquired images (Bao et al., 2023). Variations in illumination, target occlusion, and complex background interference commonly occur in field environments, requiring models with strong robustness and generalization ability (Leng et al., 2024).(3) Model deployment is constrained by computational resources (Badgujar et al., 2024). Most state-of-the-art deep learning models have a large number of parameters and high computational costs, which hinder their deployment on resource-limited agricultural devices and restrict their practical adoption in farming systems (Shehzadi et al., 2023).

In agricultural disease detection, YOLO (Redmon et al., 2016)-based models have become the mainstream baseline due to their lightweight structure and feasibility for edge deployment. Several improved variants have been proposed to address crop-specific challenges. (Gao et al., 2025) introduced SCS-YOLO for Fusarium Head Blight detection

⋆This document is the results of the research project funded by the National Science Foundation.

⋆⋆The second title footnote which is a longer text matter to fill through the whole text width and overflow into another line in the footnotes area of the first page.

This note has no numbers. In this work we demonstrate a_b the formation Y_1 of a new type of polariton on the interface between a cuprous oxide slab and a polystyrene micro-sphere placed on the slab.

*Corresponding author
**Principal corresponding author
cvr_1@tug.org.in (C. Radhakrishnan); cvr3@sayahna.org (C. Rajagopal); rishi@stmdocs.in (R. T.)
www.cvr.cc, cvr@sayahna.org (C. Radhakrishnan); www.sayahna.org (C. Rajagopal); www.stmdocs.in (R. T.)

[1]This is the first author footnote. but is common to third author as well.

[2]Another author footnote, this is a very long footnote and it should be a really long footnote. But this footnote is not yet sufficiently long enough to make two lines of footnote text.

in wheat, integrating StarNet, a CB module, and Shape-NWD loss into YOLOv5s. The model achieved 90.51% mAP with reduced parameters and was deployed on Jetson Nano, though its performance in complex backgrounds and cross-variety scenarios remains limited. (J et al., 2025) combined DenseNet with YOLO for automatic weed detection, where DenseNet reached 96% classification accuracy and YOLO achieved 92% mAP in real-time detection. This system reduces herbicide reliance and supports automation, but its dataset diversity and hardware demands limit broader application. (You et al., 2025) proposed VBP-YOLO-prune for apple detection, enhancing YOLOv8n with pruning, BiFPN, and improved loss functions. It achieved 89.0% mAP50 and 66.26% mAP50-95 with only 0.61M parameters and 3.2 GFLOPs, outperforming YOLOv8n and enabling deployment on Jetson Orin Nano. Nevertheless, its robustness in real orchard environments and multi-task scalability still require improvement. (Yao et al., 2025) developed WTAD-YOLO, the first YOLO variant integrating wavelet convolution for tomato disease detection. By introducing C3k2_WTConv, ADown, and DySample, it improved mAP@0.5 (+1.9%), recall (+3.5%), and F1 (+2.0%), while reducing parameters by 0.26M. Domain-shift experiments also confirmed strong generalization (ΔmAP@0.5 = –0.007). However, its adaptability to other crops, unseen diseases, and healthy leaf recognition remains a challenge. Taken together, these studies demonstrate the effectiveness of YOLO variants in agricultural scenarios, but also reveal unresolved issues such as cross-variety generalization, multi-task integration, and robustness under diverse and complex field conditions, which should be prioritized in future research.

To address the numerous challenges in current crop disease detection, the primary task is to enhance the model's ability to discriminate fine-grained features. We observed that Mamba (Gu and Dao, 2024) demonstrates significant advantages in this task. Its long-range dependency modeling can effectively integrate contextual information across regions, helping the model accurately differentiate between various crops and their disease types from a global perspective. Moreover, Mamba's potential extends further, as its efficient state-space model reduces the complexity of training and inference to linear scale. The YOLO series, being low-computation and highly efficient detection models, is widely adopted in crop disease detection. Therefore, we selected Mamba-YOLO (Wang et al., 2024b) as our baseline model and performed lightweight and efficiency-oriented modifications, enabling more precise detection across multi-crop and multi-disease tasks. Our improvements to the Mamba-YOLO model in terms of lightweight design and efficiency are as follows: Firstly, we found that the backbone of Mamba-YOLO exhibited limitations in feature extraction under complex tasks. To address this, we introduced a more efficient and lightweight MobileMamba (He et al., 2024) module, which enhances feature extraction while maintaining model compactness. This module integrates wavelet convolution (Finder et al., 2024) with Mamba, allowing it to capture global information while emphasizing local edge details. Furthermore, the MK-DeConv within MobileMamba employs multi-kernel depthwise separable convolution to achieve multi-receptive-field interaction, improving the model's sensitivity to lesions of various shapes. However, experiments revealed that this module still had limitations in spatial modeling and nonlinear representation. Therefore, we replaced the original depthwise separable convolution (DWConv) (Chollet, 2017) and FeedForward (Vaswani et al., 2023) structures at the end of MobileMamba with an RGBlock module based on the MLP architecture. This module improves spatial modeling and nonlinear expression while maintaining lightweight characteristics. In the neck of the network, we innovatively proposed a dual-path fusion scanning mechanism, combining SSM (S6) (Liu et al., 2024) with a serpentine scanning strategy to replace the conventional SSM (S6) scanning. Specifically, traditional SSM (S6) uses regular spatial flattening to model global path dependencies, which is suitable for capturing long-range feature correlations. The serpentine scanning simulates continuous spatial paths in images, preserving stronger local structural consistency and facilitating the extraction of fine-grained patterns and natural structural features. These two scanning methods are complementary in terms of modeling perspectives and spatial structure preservation. By fusing them, we can construct feature sequences from multiple arrangements and spatial perspectives, significantly enhancing the network's spatial representation and natural structure modeling in complex field tasks, which we refer to as SN-XSSBlock. Finally, we compared our proposed model with the latest YOLOv11. Experiments show that the C2PSA self-attention mechanism and efficient detection head in YOLOv11 provide significant advantages in inference efficiency, model lightweight design, and feature selection. Drawing on this design, we integrated similar ideas into our model. Ultimately, the model achieves high precision while improving inference speed and resource efficiency. Overall, our model demonstrates excellent performance in feature extraction, inference efficiency, parameter control, and generalization, showing strong potential for practical deployment and application.

This study focuses on leveraging the improved Mamba-YOLO model to more accurately detect multiple crop diseases in complex field environments, while maintaining lightweight parameters and computational efficiency.In summary, our main contributions are as follows:

(1)We designed a novel feature extraction module to replace the VSSBlock in the backbone of Mamba-YOLO, addressing potential feature similarity in multi-crop and multi-disease detection tasks. The new module is based on MobileMamba and incorporates an RGBlock with an MLP architecture to replace the depthwise separable convolution (DWConv) and FeedForward network at the end of MobileMamba, enhancing spatial modeling and nonlinear representation capabilities.

(2)We innovatively combined the traditional row-column scanning SSM(S6) with the proposed serpentine scanning to construct a novel SN-XSSBlock module. While the conventional row-column SSM(S6) aids in global feature

modeling, the serpentine scanning preserves the spatial continuity of natural structures, facilitating the extraction of structural features of crops. The fusion of these two scanning strategies allows feature sequences to be constructed from different orders and perspectives, significantly improving the model's spatial representation and crop structure modeling capabilities.

(3)We removed redundant detection heads in Mamba-YOLO and replaced them with the efficient and lightweight detection head from YOLOv11. Additionally, the C2PSA self-attention mechanism was introduced to select more effective features within the model.

2. Materials and methods

2.1. Prepare dataset

2.1.1. Dataset description

The dataset used in this study is named Field Plant (Moupojou et al., 2023) and is currently publicly available on Kaggle. The dataset comprises three crop species—cassava, corn, and tomato—covering a total of 24 plant diseases and 3 healthy states. Specifically, cassava includes four diseases: Cassava Bacterial Blight, Cassava Brown Leaf Spot, Cassava Mosaic, and Cassava Root Rot, as well as the healthy state (Cassava Healthy). Corn includes fifteen diseases: Corn Brown Spots, Corn Charcoal, Corn Chlorotic Leaf Spot, Corn Gray Leaf Spot, Corn Insects Damages, Corn Mildew, Corn Purple Discoloration, Corn Smut, Corn Streak, Corn Stripe, Corn Violet Decoloration, Corn Yellow Spots, Corn Yellowing, Corn Leaf Blight, and Corn Rust Leaf. Tomato includes five diseases: Tomato Brown Spots, Tomato Bacterial Wilt, Tomato Blight Leaf, Tomato Leaf Mosaic Virus, and Tomato Leaf Yellow Virus. Figure 1 illustrates a subset of the diseases in the dataset.

2.1.2. Data augmentation

We found that the dataset suffers from significant class imbalance, with many categories having relatively few samples. To address this, we employed Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) for data augmentation. Traditional augmentation methods typically involve image flipping, cropping, translation, or adding noise, whereas GANs are capable of generating diverse and stylistically varied images, thereby enhancing the model's generalization ability. To better reflect real-world farmland conditions, we applied augmentation conservatively: only categories with fewer samples were augmented using GANs, and the image styles were not excessively altered. Each disease image was augmented with five additional images without changing the relative position of the disease, and the original labels were transferred to the new images. Our augmentation strategy ensures that GAN-generated images maintain a high degree of structural similarity to the original ones, while introducing slight variations in local textures, details, and pixel distributions. This allows the model to better adapt to minor perturbations. The augmentation results are illustrated in Figure 2.

We employed the Structural Similarity Index Measure (SSIM) (Wang et al., 2004) to analyze image inconsistencies by calculating the SSIM scores between the original image and each of the five augmented images. The results are as follows:SSIM between image1 and image2: 0.8410.SSIM between image1 and image3: 0.8406.SSIM between image1 and image4: 0.8412.SSIM between image1 and image5: 0.8407.SSIM between image1 and image6: 0.8415.This augmentation strategy for minority classes improved the model's ability to memorize samples from underrepresented categories, effectively alleviating the class imbalance issue and reducing the risk of overfitting.Finally, we compared the sample distributions before and after augmentation, as shown in Figure 3. The final augmented dataset was split into training, validation, and testing sets with a ratio of 7:2:1, resulting in 5,075 images for training, 1,450 for validation, and 726 for testing.

2.2. MSN-Mamba-YOLO

The MSN-Mamba-YOLO object detection model is designed based on Mamba-YOLO, featuring strong feature extraction capabilities and excellent environmental adaptability.The architecture of MSN-Mamba-YOLO is shown in Figure 4.

In the backbone, we designed a lightweight RG-MMamba module based on the MobileMamba block to replace the original ODSSBlock, enhancing feature representation in complex field scenarios and reducing feature similarity. Additionally, the CSPSA attention mechanism from YOLOv11 is introduced at the end of the backbone to improve the recognition of fine-grained features, maintaining clear feature distinctions for multi-scale and multi-morphology disease targets. We also retain the Simple Stem and Vision Clue Merge modules: the former achieves a better balance between performance and efficiency, while the latter fuses multi-level features to effectively integrate critical information and reduce missed detections for large, medium, and small disease targets. In the neck, we follow Mamba-YOLO's PAFPN (Liu et al., 2018)-based multi-scale feature fusion design and propose a parallel dual-scan modeling strategy: retaining the original row-column SSM while introducing a novel serpentine scan mechanism. The fused outputs capture both contextual dependencies and structural information, enhancing feature interaction and semantic understanding. This module, named SN-XSSBlock, replaces the ODSSBlock in the neck. For the detection head, we streamline the original design by removing redundant heads and adopting YOLOv11's lightweight and efficient detection head, achieving a synergistic improvement in inference speed and detection performance.

2.2.1. backbone part

The backbone network is primarily responsible for feature extraction. To enhance the model's discriminative capability during feature extraction, we observed that MobileMamba can efficiently extract informative features, but its nonlinear representation and spatial modeling abilities are somewhat limited. To address this, we introduce the RGBlock to replace the DW-Conv and FeedForwardNet at the end of

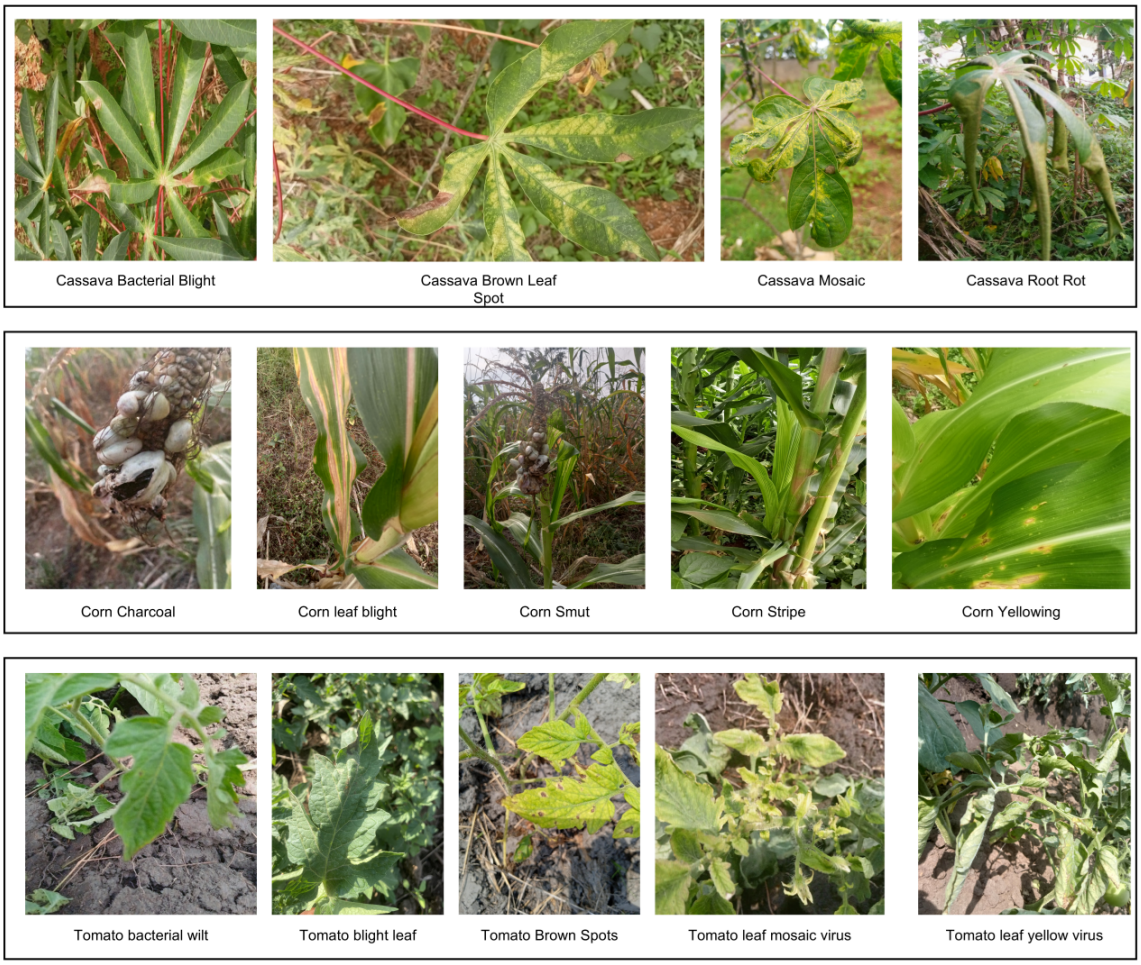

Fig. 1. some diseases in the Field Plant dataset

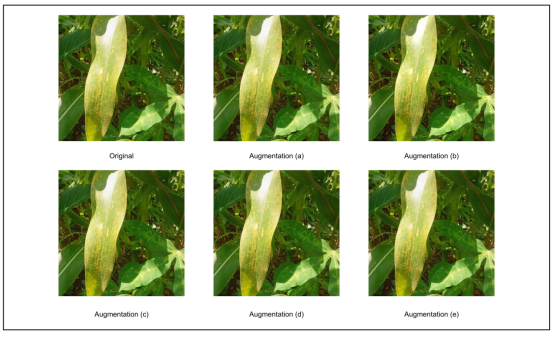

Fig. 2. Example of GAN-based data augmentation

MobileMamba, improving nonlinear representation and spatial modeling while effectively reducing parameters and computational cost. The structure of the RG-MMamba model is shown in Figure 5.

The RG-MMamba module mainly consists of three parts. The first part uses DW-Conv and FeedForwardNet to reduce feature redundancy and learn richer features.The second part employs MRFFI (Multi-Receptive Field Feature Interaction module) to achieve efficient interaction among features from multiple receptive fields. The third part leverages the RGBlock to improve the model's nonlinear capacity.Within MRFFI, the feature $x^O \in \mathbb{R}^{h\times w\times c}$ is divided into three parts. The first part is the Long-range WTE-Mamba, which excels at feature extraction. This module simultaneously performs global and local learning on the feature block $x_G^O \in \mathbb{R}^{h\times w\times \xi c}$, where $0<\xi<1$. By passing this feature block through the bidirectional scanning Mamba module, the model can more effectively capture relationships between distant pixels in the image, enabling it to understand the contextual relationships of plant diseases in the image and compensating for the limited receptive field of traditional convolutions.

$$\begin{aligned} x_{m1}^O &= SS2D\Big(\sigma\big(Conv\big(Linear(x_G^O)[:\,\xi c]\big)\big)\Big),\\ x_{m2}^O &= \sigma\big(Linear(x_G^O)[\xi c\,:]\big),\\ x_m^I &= Linear\big(x_{m1}^O \otimes x_{m2}^O\big). \end{aligned} \tag{1}$$

At the same time, Haar wavelet transform is applied to the same feature block for multi-scale frequency-domain decomposition, obtaining feature representations $x_G^O \in \mathbb{R}^{\frac{h}{2}\times\frac{w}{2}\times 4\xi c}$ different frequency bands. Local convolutions

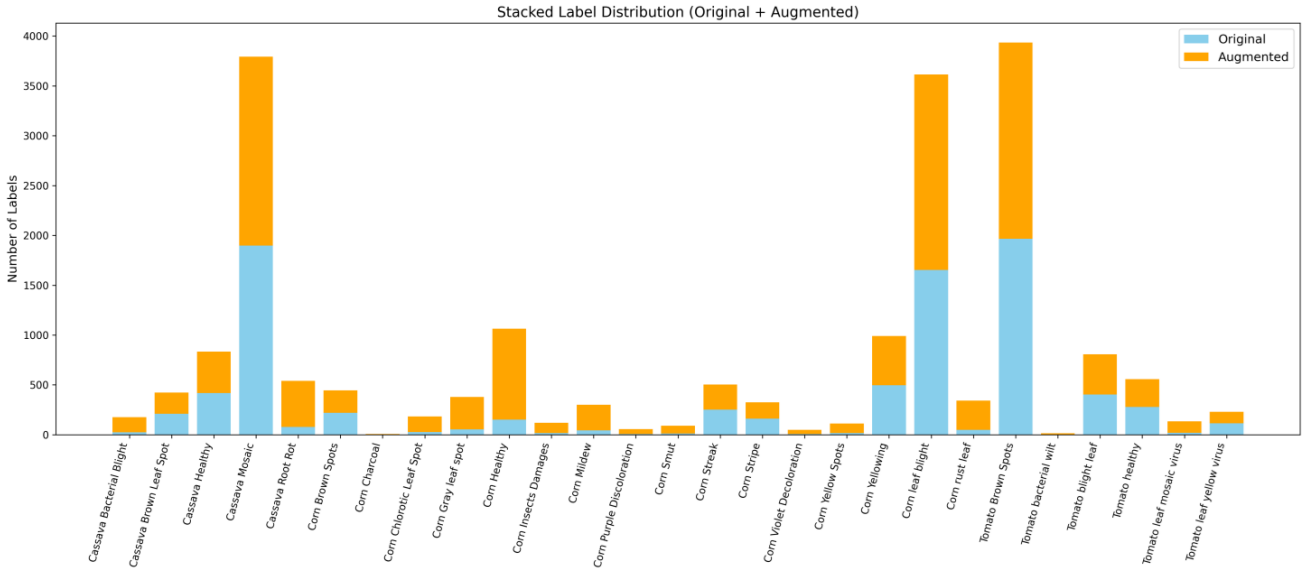

Fig. 3. Example of GAN-based data augmentation

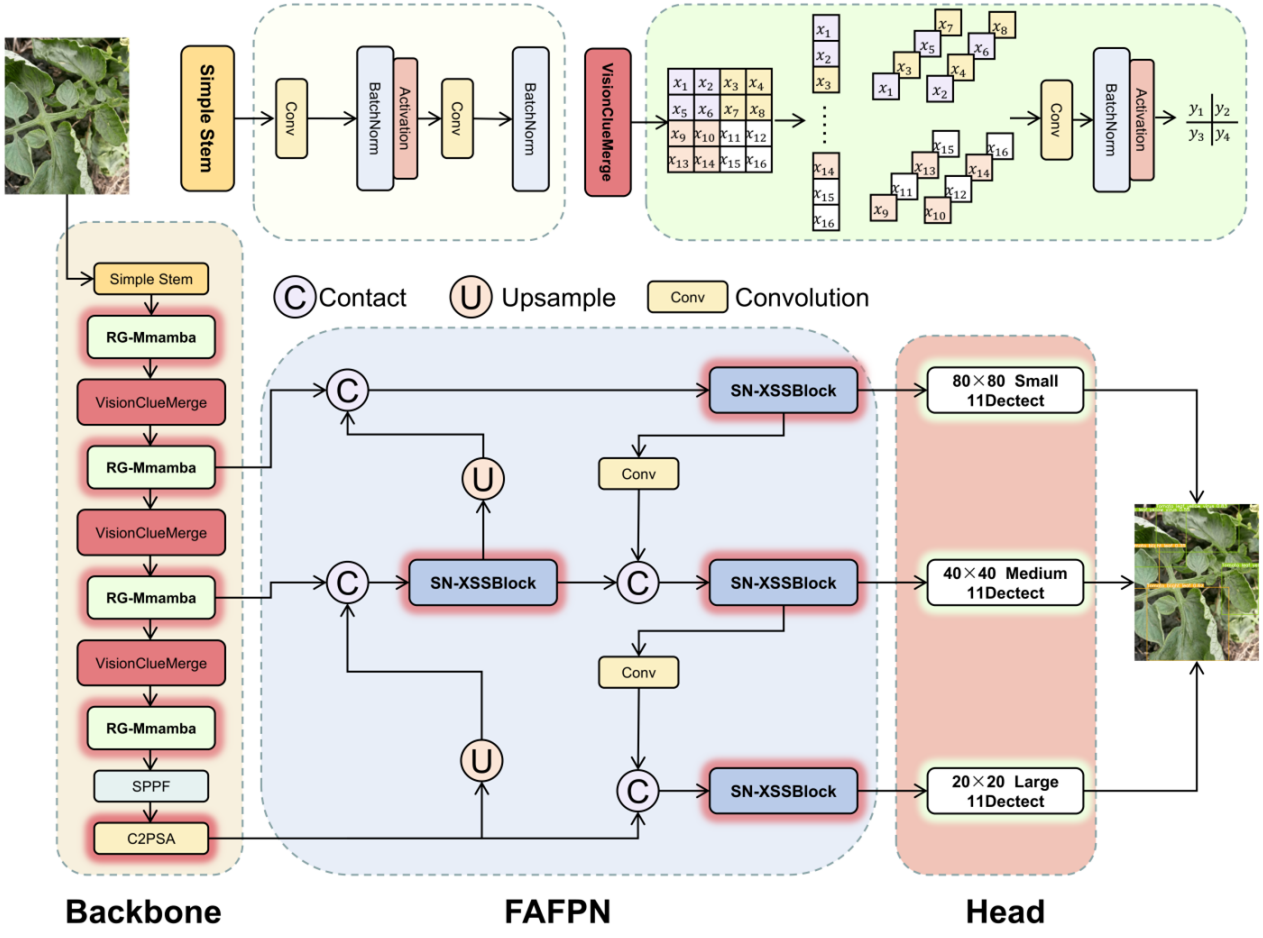

Fig. 4. MSN-Mamba-YOLO structure diagram

are then performed to extract information, capturing high-frequency edge details and fine textures in the image.Finally, inverse wavelet transform is applied to restore the features to

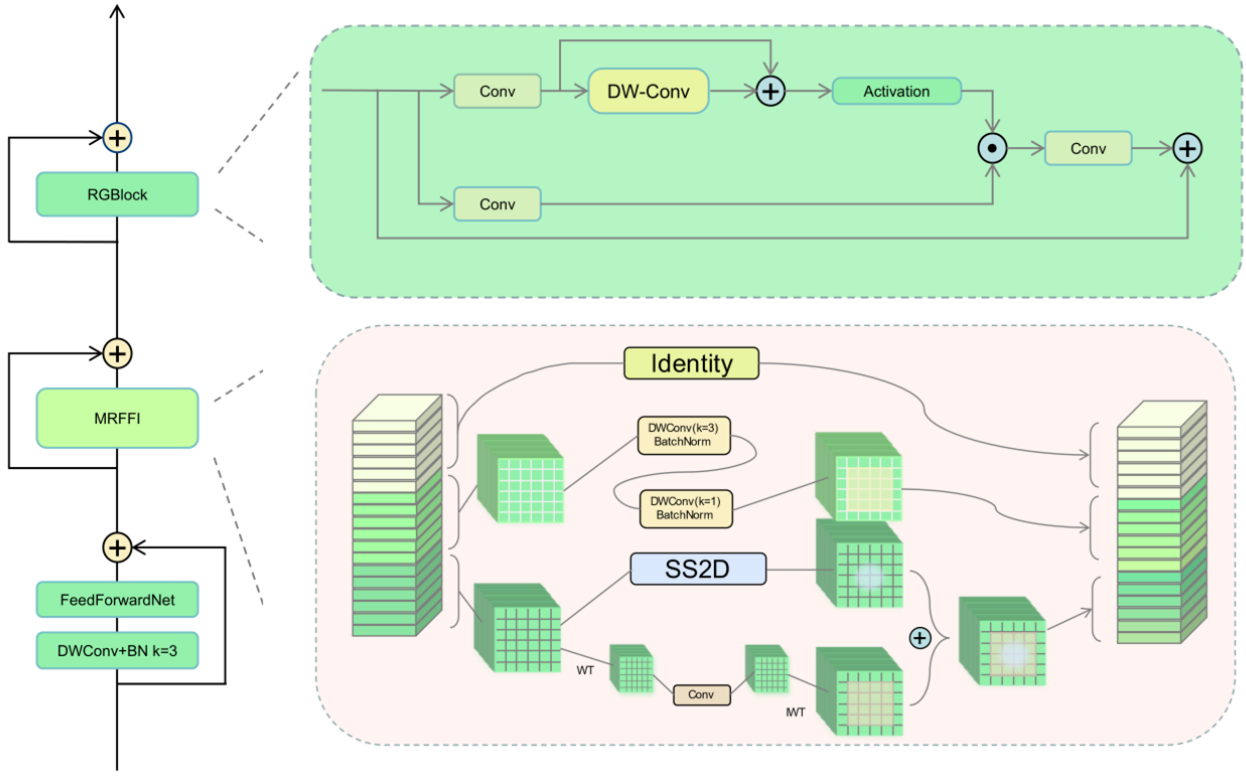

Fig. 5. RG-MMamba structure diagram

the original channel mapping size $x_U^I \in \mathbb{R}^{h\times w\times \xi c}$.

$$\begin{aligned} x_{wt}^O &= WT\Big(x_U^O, \big[f_{LL}, f_{LH}, f_{HL}, f_{HH}\big]\Big), \\ x_U^I &= IWT\Big(Conv(x_{wt}^O), \big[f_{LL}, f_{LH}, f_{HL}, f_{HH}\big]\Big). \end{aligned} \tag{2}$$

Here, $f_{LL} = \frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$ corresponds to the low-pass filter, while $f_{LH} = \frac{1}{2}\begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix}$, $f_{HL} = \frac{1}{2}\begin{bmatrix}1 & 1\\ -1 & -1\end{bmatrix}$, and $f_{HH} = \frac{1}{2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$ form a set of high-pass filters. The final output of this part is obtained by adding the wavelet-transformed features to the output feature map of the Mamba module. This design not only strengthens the model's capacity for contextual understanding but also makes it more sensitive to regions with significant image variations (high-frequency information). In scenarios involving multiple crops and multiple diseases, this design enables the extracted features to achieve mutual understanding between global and local information.

$$x_G^I = x_m^I + x_U^I, \quad \text{where} \quad x_G^I \in \mathbb{R}^{h\times w\times \xi c} \tag{3}$$

For the second part of the features $x_L^O \in \mathbb{R}^{h\times w\times \mu c}$, where $0<\mu<1$, the channels are divided into $n \in \mathbb{N}$ groups, each denoted as $x_{Lj}^O \in \mathbb{R}^{h\times w\times \frac{\mu c}{n}}$. Different convolution kernels of various sizes are applied to each group, and the outputs of these convolutions are concatenated to obtain the final output feature $x_L^I \in \mathbb{R}^{h\times w\times \mu c}$. This design augments the potential to capture lesion characteristics on crop leaves with diverse shapes and sizes, ranging from tiny spots to large-scale infections.For the third part of the features, we directly map the remaining $(1-\xi-\mu)c$ features, retaining a certain amount of the original features to reduce redundancy in high-dimensional intermediate features and significantly improve computational efficiency. Finally, the processed features obtained through the MRFFI module are denoted as $x^I \in \mathbb{R}^{h\times w\times c}$.

$$\begin{aligned} x_{Lj}^I &= \mathrm{Conv}\big(x_{Lj}^I, k=(2j+1)\big), \quad j=1,\dots,n, \\ x_L^I &= \mathrm{Concat}\Big([x_{L1}^O,\dots,x_{Ln}^O], \dim=-1\Big), \\ x^I &= \mathrm{Concat}\Big(x_G^I, x_L^I, x^O[(1-\xi-\mu)c:]\Big). \end{aligned} \tag{4}$$

Finally, the processed feature block $x^I \in \mathbb{R}^{h\times w\times c}$ is fed into the RG Block module. This module uses a gating mechanism to split the input features into two paths: local information $x_{K1}^I \in \mathbb{R}^{h\times w\times \frac{c}{2}}$ and global information $x_{K2}^I \in \mathbb{R}^{h\times w\times \frac{c}{2}}$. In the global path, a depthwise convolution (DW-Conv) is employed as a positional encoding module, and training is performed using residual cascades to facilitate more effective gradient flow and reduce computational cost. Here,Φ denotes the GELU activation function. The processed global features are then multiplied with the local features, and

the result is added to the original features to obtain the final channel representation $x^P \in \mathbb{R}^{h\times w\times c}$.

$$\begin{aligned} x_{K1}^I, x_{K2}^I &= \text{chunk}\Big(Conv(x^I), k=1\Big), \\ x_{K2}^T &= \Phi\big(DWConv(x_{K2}^I) \oplus x_{K2}^I\big), \\ x^P &= Conv\big(x_{K2}^T \otimes x_{K1}^I\big) + x^I. \end{aligned} \tag{5}$$

This module effectively integrates fine-grained lesion features with the overall morphology of diseased leaves, significantly improving the model's discriminative capability when dealing with complex lesion patterns, varying lesion scales, and background interference. Furthermore, by combining depthwise separable convolutions with a residual concatenation design, the module preserves the spatial structure of the image and facilitates efficient gradient propagation, while maintaining low computational overhead, making it highly lightweight.In summary, the RG Block boosts the model's potential to capture spatial details and improves feature representation, thereby increasing the robustness and performance of the disease detection model under conditions involving multiple disease types, small lesions, and complex field backgrounds. At the end of the backbone network, we incorporate the C2PSA module from the YOLO11 series, whose structure is shown in Figure 6.

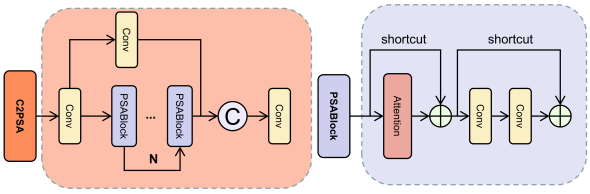

Fig. 6. C2PSA structural diagram:This module is based on the cross-stage partial (CSP) splitting strategy, dividing the input features into a main branch and a side branch. The side branch employs compressed convolutions and spatial attention modeling to emphasize target regions and key textures, while the main branch maintains information flow to reduce redundant computations. Finally, channel-wise concatenation and fusion convolutions are used to achieve information feedback and scale alignment, thereby selecting and enhancing more discriminative features with minimal additional inference cost.

2.2.2. Neck part

In the neck of the model, we employ a Dual-Path Concurrent Global Scanning Mechanism to process the fused features. This module is designed to model global semantic information of the image while effectively preserving the structural characteristics of plant leaves. The mechanism is integrated into the proposed SN-XSSBlock module. The structural diagram of the model is shown in Figure 7.

First, the input feature $x^E \in \mathbb{R}^{h\times w\times c}$ is processed by the local feature extraction module LSBlock, which leverages depthwise separable convolution and batch normalization to model specific local spatial features, thereby enabling the model to better capture image details. The resulting feature is denoted as $x^Q \in \mathbb{R}^{h\times w\times c}$. Subsequently, the feature $x^Q \in \mathbb{R}^{h\times w\times c}$ is fed into the core module SN-SS2D, where, after being processed by a depthwise convolution (DW-Conv), the channels are divided into two parallel paths, yielding the feature representation $x_{i1}^Y, x_{i2}^Y \in \mathbb{R}^{h\times w\times \frac{c}{2}}$.

$$\begin{aligned} x^R &= BN\Big(DWConv(x^E)\Big), \\ x^Q &= Conv\Big(\sigma\big(Conv(x^R)\big)\Big), \\ x_{i1}^Y, x_{i2}^Y &= \text{chunk}\Big(\sigma\big(DWConv(x^Q)\big)\Big). \end{aligned} \tag{6}$$

The first path is the traditional row–column scanning path S6, which enhances the model's global receptive field. The second path is the serpentine scanning path, designed to preserve the continuity and integrity of leaf structures, thereby improving the model's sensitivity to target shapes. After parallel modeling, the two paths are integrated through a fusion mechanism, and residual connections are introduced to facilitate feature flow and ensure network stability. Finally, the fused features are fed into the RGBlock module, which provides spatial detail modeling, yielding the final feature output $x^C \in \mathbb{R}^{h\times w\times c}$.

$$\begin{aligned} x^N &= Scan(x_{i1}^Y) \oplus Scan(x_{i2}^Y), \\ x^Z &= Linear\Big(LN(x_i^N) \otimes x^E\Big), \\ x^C &= RGBlock\Big(LN(x^Z)\Big) \oplus x^Z. \end{aligned} \tag{7}$$

The traditional row–column scanning method S6 effectively models long-range dependencies in both horizontal and vertical directions, facilitating the capture of cross-regional semantic information. In contrast, the serpentine scanning approach employs continuous and alternating scanning paths to preserve spatial coherence and local consistency, with a stronger focus on the natural spatial distribution of objects in images. This is particularly advantageous for modeling the overall shape of leaves, edge contours, and the directional spread of lesions along leaf veins. The fusion of the two methods enables collaborative modeling of both local and global, The fusion of the two methods enables the collaborative modeling of local and global information, as well as structural and semantic information, thereby enhancing the model's disease detection competence in complex real-world field scenarios. Overall, the design of this module improves the model's structural modeling proficiency in complex agricultural scenarios.

2.2.3. Head Part

The detection head used in Mamba-YOLO is derived from YOLOv8. We found that this detection head is not only structurally redundant but also exhibits limited detection capability. Therefore, we replaced it with a more lightweight detection head from YOLO11. Compared to the original Mamba-YOLO detection head, this head incorporates two

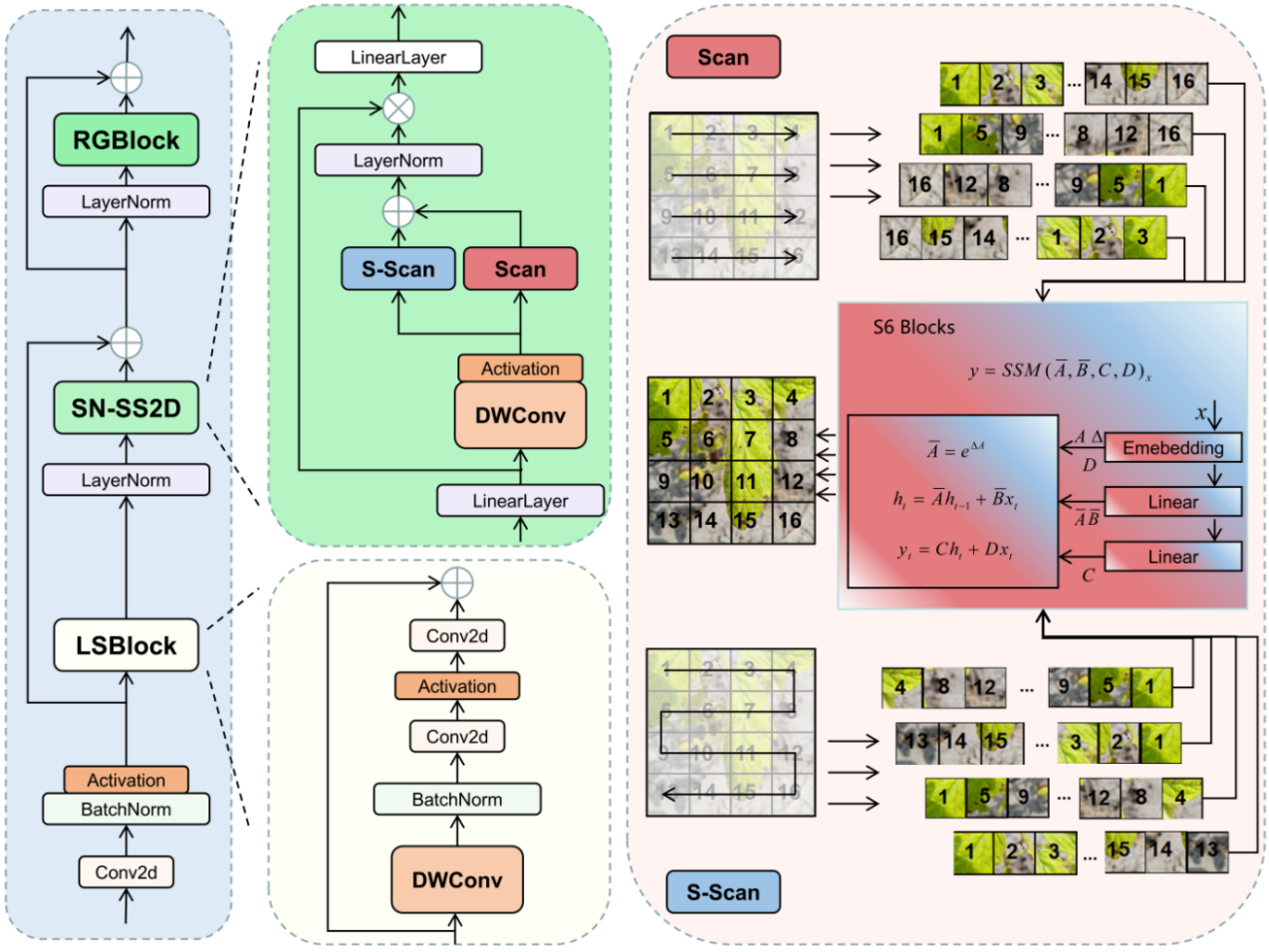

Fig. 7. SN-XSSBlock structural diagram

depthwise separable convolutions (DW-Conv) along with a 1×1 convolution, which reduces computational cost while maintaining model performance. The structure of this detection head is shown in Figure 8.

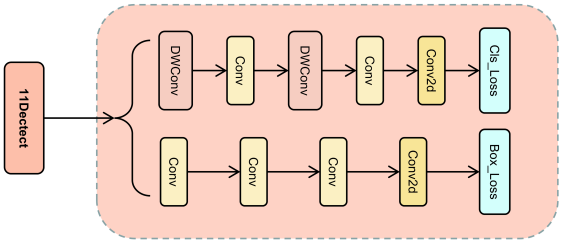

Fig. 8. YOLO11Dectect

3. Results and discussion

3.1. Experimental Environment

Table 1 presents the detailed parameters of our experimental setup. We adopt the SGD optimization strategy, with the input image resized uniformly to 640×640. The learning rate is set to 0.001, four subprocesses are used for parallel data loading, and a batch size of 12 is employed for efficient training.

Table 1
Experimental Configuration Parameters

Items	Types
Operating system	Ubuntu 20.04
CPU	12th Gen Intel(R) Core(TM) i5-12600KF
GPU	RTX 4070 SUPER (12 GB)
Language	Python 3.10
Accelerating environment	CUDA 11.8
Framework	Torch 2.1

3.2. Experimental Evaluation Metrics

To evaluate the effectiveness of the proposed multi-class, multi-disease detection method, this study adopts commonly used metrics in deep learning, including Precision, Recall, mean Average Precision (mAP), Giga Floating Point Operations per second (GFLOPs), and the number of

model parameters.Precision represents the proportion of true positive samples among all samples predicted as positive by the model; Recall reflects the proportion of actual positive samples correctly identified by the model. Average Precision (AP) quantifies the trade-off between Precision and Recall, typically measured as the area under the Precision-Recall curve. mAP (mean Average Precision) averages the AP across all classes to assess overall detection performance, where higher values indicate better performance. Specifically, mAP@0.5 is computed at an IoU threshold of 0.5, while mAP@50:95 averages results over IoU thresholds from 0.5 to 0.95 with a step size of 0.05, providing a more comprehensive evaluation under varying matching conditions. The F1-score integrates both Precision and Recall, offering an overall measure of the model's predictive capability. GFLOPs quantify the number of floating-point operations required for a single forward inference, where smaller values indicate higher computational efficiency.Considering all metrics, this study adopts mAP@50:95 as the core evaluation criterion to measure the model's performance in complex multi-class plant disease detection tasks.

$$
\begin{aligned}
P &= \frac{TP}{TP + FP}, \\
R &= \frac{TP}{TP + FN}, \\
F1 &= \frac{2TP}{2TP + FP + FN}, \\
mAP_{0.5} &= \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i^{\mathrm{IoU}=0.5}, \\
mAP_{[0.5:0.95]} &= \frac{1}{10}\sum_{k=0}^{9}\left(\frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i^{\mathrm{IoU}=0.5+0.05k}\right).
\end{aligned}
\tag{8}
$$

3.3. Experimental Evaluation Metrics

Table 2 presents a series of ablation experiments based on the Mamba-YOLO-T baseline model, aiming to address three core challenges in multi-class crop disease detection—feature similarity, complex scene modeling, and model redundancy—through step-by-step structural optimization. On the enhanced Field Plant dataset, our model achieved an mAP@50–95 of 80.5%, with an average inference time of only 3.7 ms per image. Compared with the baseline Mamba-YOLO-T, inference speed was improved by approximately 2×, mAP increased by 0.7%, the number of parameters was reduced by 22%, and computational cost decreased by 30%. These results indicate that, while maintaining high detection accuracy, our model significantly enhances computational efficiency and real-time inference capability, demonstrating strong practicality and deployment potential.

Figure 9 illustrates the comparison of Loss and mAP@50:95 trends between the improved model and the baseline model on the validation set. As shown by the curves, both models exhibit a generally consistent downward trend in Loss during training. However, in terms of accuracy, the improved model begins to demonstrate an advantage after the 50th epoch and continues to maintain its lead. Although the two models reach comparable Loss levels, our model achieves superior detection accuracy, clearly demonstrating the effectiveness and superiority of its architectural design.

A. My Appendix

Appendix sections are coded under `\appendix`.

`\printcredits` command is used after appendix sections to list author credit taxonomy contribution roles tagged using `\credit` in frontmatter.

CRediT authorship contribution statement

CV Radhakrishnan: Conceptualization of this study, Methodology, Software. **CV Rajagopal:** Data curation, Writing - Original draft preparation.

References

Badgujar, C.M., Poulose, A., Gan, H., 2024. Agricultural object detection with you only look once (yolo) algorithm: A bibliometric and systematic literature review. Computers and Electronics in Agriculture 223, 109090. URL: https://www.sciencedirect.com/science/article/pii/S0168169924004812, doi:https://doi.org/10.1016/j.compag.2024.109090.

Bangal, S.P., Jondhale, S.R., Agarkar, B.S., Chaudhari, S.V., Shaikh, S.A., 2025. Recent advances, challenges and future prospects of multi-crop leaf disease detection algorithm using deep learning: A comprehensive review, in: 2025 International Conference on Inventive Computation Technologies (ICICT), pp. 246–251. doi:10.1109/ICICT64420.2025.11004667.

Bao, W., Zhu, Z., Hu, G., Zhou, X., Zhang, D., Yang, X., 2023. Uav remote sensing detection of tea leaf blight based on ddma-yolo. Computers and Electronics in Agriculture 205, 107637. URL: https://www.sciencedirect.com/science/article/pii/S016816992300025X, doi:https://doi.org/10.1016/j.compag.2023.107637.

Barbedo, J.G.A., 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Computers and Electronics in Agriculture 153, 46–53. URL: https://www.sciencedirect.com/science/article/pii/S0168169918304617, doi:https://doi.org/10.1016/j.compag.2018.08.013.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. URL: https://arxiv.org/abs/1610.02357, arXiv:1610.02357.

Finder, S.E., Amoyal, R., Treister, E., Freifeld, O., 2024. Wavelet convolutions for large receptive fields. URL: https://arxiv.org/abs/2407.05848, arXiv:2407.05848.

Gao, C., He, B., Guo, W., Qu, Y., Wang, Q., Dong, W., 2025. Scs-yolo: A real-time detection model for agricultural diseases — a case study of wheat fusarium head blight. Computers and Electronics in Agriculture 238, 110794. URL: https://www.sciencedirect.com/science/article/pii/S0168169925009007, doi:https://doi.org/10.1016/j.compag.2025.110794.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. URL: https://arxiv.org/abs/1406.2661, arXiv:1406.2661.

Gu, A., Dao, T., 2024. Mamba: Linear-time sequence modeling with selective state spaces. URL: https://arxiv.org/abs/2312.00752, arXiv:2312.00752.

He, H., Zhang, J., Cai, Y., Chen, H., Hu, X., Gan, Z., Wang, Y., Wang, C., Wu, Y., Xie, L., 2024. Mobilemamba: Lightweight multi-receptive visual mamba network. URL: https://arxiv.org/abs/2411.15941, arXiv:2411.15941.

J, J., Abhiram, K., Abhishek, C., Sharma, D.A., 2025. An automated weed-detection approach using deep learning in agriculture system, in: 2025 International Conference on Advanced Computing Technologies (ICoACT), pp. 1–7. doi:10.1109/ICoACT63339.2025.11004820.

Leng, X., Chen, J., Huang, J., Zhang, L., Li, Z., 2024. An improved yolov8-based method for real-time detection of harmful tea leaves in complex backgrounds. Phyton-International Journal of Experimental Botany

Table 2
Ablation Experiments (Validation Set)

Variant	Mamba-yolo-T	MobileMamba	RG-MMamba	SN-XSSBlock	v11Detect / C2PSA	GFLOPs	Params (MB)	mAP50-95 (%)
(a)	✓	-	-	-	-	13.6	5.99	79.8
(b)	✓	✓	-	-	-	11.2	4.65	79.6
(c)	✓	✓	-	-	✓	10.3	4.61	80.1
(d)	✓	✓	✓	-	-	10.5	4.89	80.0
(e)	✓	✓	✓	✓	-	10.5	4.72	80.2
(f)	✓	✓	✓	-	✓	9.6	4.86	80.4
(g)	✓	✓	✓	✓	✓	9.6	4.68	80.5

93, 2963–2981. URL: http://www.techscience.com/phyton/v93n11/58785, doi:10.32604/phyton.2024.057166.

Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path aggregation network for instance segmentation. URL: https://arxiv.org/abs/1803.01534, arXiv:1803.01534.

Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y., 2024. Vmamba: Visual state space model. URL: https://arxiv.org/abs/2401.10166, arXiv:2401.10166.

Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., Tapamo, H., Nkenlifack, M., 2023. Fieldplant: A dataset of field plant images for plant disease detection and classification with deep learning. IEEE Access 11, 35398–35410.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788.

Rezk, N.G., Attia, A.F., El-Rashidy, M.A., et al., 2022. An efficient plant disease recognition system using hybrid convolutional neural networks (cnns) and conditional random fields (crfs) for smart iot applications in agriculture. International Journal of Computational Intelligence Systems 15, 65. URL: https://doi.org/10.1007/s44196-022-00129-x, doi:10.1007/s44196-022-00129-x.

Shehzadi, T., Hashmi, K.A., Stricker, D., Afzal, M.Z., 2023. Object detection with transformers: A review. URL: https://arxiv.org/abs/2306.04670, arXiv:2306.04670.

Upadhyay, A., G C, S., Das, S., Mettler, J., Howatt, K., Sun, X., 2025. Multiclass weed and crop detection using optimized yolo models on edge devices. Journal of Agriculture and Food Research 22, 102144. URL: https://www.sciencedirect.com/science/article/pii/S2666154325005150, doi:https://doi.org/10.1016/j.jafr.2025.102144.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention is all you need. URL: https://arxiv.org/abs/1706.03762, arXiv:1706.03762.

Wang, Y., Xu, C., Wang, Y., Wang, X., Ding, W., 2024a. Adversarially attack feature similarity for fine-grained visual classification. Applied Soft Computing 163, 111945. URL: https://www.sciencedirect.com/science/article/pii/S1568494624007191, doi:https://doi.org/10.1016/j.asoc.2024.111945.

Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612. doi:10.1109/TIP.2003.819861.

Wang, Z., Li, C., Xu, H., Zhu, X., Li, H., 2024b. Mamba yolo: A simple baseline for object detection with state space model. URL: https://arxiv.org/abs/2406.05835, arXiv:2406.05835.

Yao, J., Li, Y., Xia, Z., Nie, P., Li, X., Li, Z., 2025. Wtad-yolo: A lightweight tomato leaf disease detection model based on yolo11. Smart Agricultural Technology 12, 101349. URL: https://www.sciencedirect.com/science/article/pii/S2772375525005805, doi:https://doi.org/10.1016/j.atech.2025.101349.

You, H., Wang, H., Wei, Z., Bi, C., Zhang, L., Li, X., Yin, Y., 2025. Vbp-yolo-prune: Robust apple detection under variable weather via feature-adaptive fusion and efficient yolo pruning. Alexandria Engineering Journal 128, 992–1014. URL: https://www.sciencedirect.com/science/article/pii/S1110016825008828, doi:https://doi.org/10.1016/j.aej.2025.08.013.

Author biography without author photo. Author biography.

Author biography with author photo. Author biography.

Author biography with author photo. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography. Author biography.