MPGA-YOLOv8: A Leaf Disease Detection Network Combining Multi-Scale Progressive Feature Network and Global Attention

Jiacheng Xu, Yonghui Yang and Maoxiang Chu

***Abstract*—Considering the diversity of natural conditions, including fog,low light,and strong light, as well as the impact of various diseases on leaves, we propose an improved apple leaf disease detection method based on the YOLOv8n model. This method first uses a Multi-Scale Progressive Feature Network(MSPN) as the neck network to integrate multiple morphological lesions and enhance the model's information integration capability. Then, a Global Self Attention Module (GSAM) is applied at the end of the backbone to help the model capture spatial relationships in the image and reduce interference from extremely complex background conditions. Next, we replace the convolution blocks in the backbone with Receptive Field-Focused Convolution Block (RFFconv) to effectively improve the model's recognition accuracy through shared receptive field weights. Finally, we add a small object detection layer for small lesions to enhance the model's generalization performance. Experimental results show that the proposed MPGA-YOLOv8 model effectively detects six types of apple leaf diseases in complex backgrounds, including healthy samples, with an average mAP accuracy of 74.3%. Compared to classic models like SSD, Faster RCNN, YOLOv3 tiny, YOLOv5, and YOLOv6, the mAP performance improves by 20.9%, 38.4%, 9.8%, 10.0%, and 14.3%, respectively. This model enables rapid and accurate detection and recognition of apple leaf diseases, providing viable technical support and solutions for disease prevention and control.**

***Index Terms*—Appleleafdisease detection, Attention mechanism, Field-Focused Convolution, Multi-Scale feature.**

I. Introduction

APPLE as one of the widely grown fruits in the world, requires a significant investment of manpower and resources every year to cultivate and cultivate apples, in order to ensure their quality and yield. However, monitoring and controlling apple leaf diseases during the growth process is crucial [1]. The traditional manual monitoring method is not only inefficient, but also has strong subjectivity and high misjudgment rate, which further increases the cost of artificial planting and the effect is not satisfactory.

In the 1980s, scholars began using advanced digital image processing and feature egineering for diagnosing plant diseases, with research evolving from early image preprocessing to feature extraction, disease spot segmentation, and detection [2]. Image processing techniques include edge detection, color segmentation, and texture analysis. Feature engineering involves manual feature extraction and machine learning algorithms like support vector machines (SVM) and random forests [3]. While these methods are simple and user-friendly, they have low accuracy and robustness in complex situations. Their ability to extract nonlinear features and adapt to varied scenes is limited. Additionally, extracting features in complex backgrounds demands significant manpower and resources, hindering wider adoption.

In recent years, deep learning convolutional neural networks have addressed limitations in traditional methods. Compared to conventional vision techniques, deep learning models offer higher accuracy, flexibility, and adaptability. Object detection is a key task in computer vision, known for its real-time performance. Detection methods are mainly categorized into single-stage and two-stage detectors. Two-stage detectors first generate candidate regions, then classify and refine them [4].Single-stage object detection has seen significant research in agricultural disease detection. In 2022, Li et al.[5] proposed a multi-scale feature fusion method for detecting corn leaf diseases using convolutional neural networks (CNN). Their experiments tackled complex conditions, such as overlapping occlusions and similar textures in disease areas, providing a feasible solution for detecting corn plant diseases. Also in 2022, Chen et al.[6] introduced a cucumber leaf disease detection method based on an improved Fast Region-based CNN. Their results showed an mAP value of 83.

While two-stage detectors excel in accuracy, their complex steps and multiple modules for generating candidate regions slow down detection speed. In agricultural scenarios requiring real-time performance, single-stage detectors are more advantageous. To address the limitations of two-stage detectors in achieving real-time performance, single-stage detectors directly generate bounding boxes on images and classify them, with YOLO and SSD series as notable examples. Jiang et al.[7] proposed an object detection algorithm based on INAR-SSD for real-time detection of apple leaf diseases. By incorporating the GoogLeNet Inception structure and Rainbow cascade, they detected five common apple leaf diseases—Alternaria leaf spot, brown spot, mosaic, gray spot, and rust—achieving a detection performance of 78.80% mAP. Liu et al.[8] introduced a lightweight ShuffleNetv2 network with a CBAM attention mechanism in the YOLOv3 model, improving the accuracy of grape leaf disease and insect detection to 90.4%. The Sun team[9] enhanced YOLOv5 by adding a Ghost structure (Ghost conv

Manuscript received August 3, 2024; revised November 30, 2024. Thiswork was supported by the Special Fund for Scientific Research Construction of University of Science and Technology Liaoning, China.

Jiacheng Xu is a Postgraduate Student of School of Electronic Information, University of Science and Technology Liaoning, Anshan, 114051 China. (e-mail: 13359473587@163.com).

Yonghui Yang is a Professor of School of Control Science andEngineering, University of Science and Technology Liaoning, Anshan,114051 China. (Corresponding author to provide phone: 86-0412-5929068; e-mail:yangyh2636688@163.com).

Maoxiang Chu is a Professor of School of Control Science and Engineering, University of Science and Technology Liaoning, Anshan, 114051 China. (e-mail: chu52_2004@163.com).

and Ghost Bottleneck), CBAM attention mechanism, and a bidirectional feature pyramid network (BiFPN), achieving 90.9% accuracy for four apple leaf diseases: bitter fruit disease, anthracnose, ring disease, and fruit rust.Yue et al.[10] proposed an improved YOLOX Nano model for detecting apple leaf lesions in 2022. They enhanced the YOLOX Nano backbone network using an asymmetric ShuffleBlock, CSP-SA module, and blueprint separable convolution (BSConv), significantly boosting feature extraction and detection performance.Despite the progress of single-stage detectors in plant leaf disease detection, challenges remain, such as low detection accuracy for small diseases, limited identifiable categories, false detections from multiple diseases on leaves, and issues in noisy conditions. Considering the complexity of the apple orchard environment and the need for real-time detection, this study proposes an improved algorithm based on the YOLOv8n model to effectively address these challenges and enhance the detection speed of apple leaf diseases.

II. MATERIALS AND METHOD

A. Data acquisition and labeling

The AppleLeaf dataset is a comprehensive fusion of four distinct datasets: the renowned PlantVillage dataset, the AppleLeaf Disease Segmentation Dataset (ATLDSD), as well as the specialized PPCD2020 and PPCD2021 datasets[11]. This integration of diverse sources ensures a robust and varied dataset, providing a solid foundation for effective apple leaf disease detection and segmentation.

The PlantVillage dataset is a publicly available large-scale plant disease image dataset created and maintained by researchers at Cornell University. The dataset aims to facilitate the automatic identification and study of plant diseases and support scientific research in the field of plant health care and agriculture. Fig.1. shows the dataset.

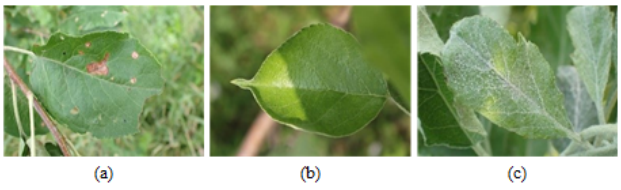

Fig. 1. AppleLeaf dataset (a) frog eye leaf spots; (b) healthy; (c) powdery mildew

AppleLeaf Disease Segmentation dataset (ATLDSD) was collected from four different apple experimental demonstration stations. ATLDSD was collected in the laboratory (about 51.9%) and in the field (about 48.1%) under different weather conditions. At the same time, because some disease categories of ATLDSD, PPCD2020, and PPCD2021 are the same, some images of the three datasets are fused. This is shown in Fig.2.

From the collected dataset, we selected five types of diseased leaves and healthy leaves for training, including frogeye leaf spot, powdery mildew, rust, scab, and brown spot, along with six types of healthy leaves. To enhance sample diversity, we included images taken under varying lighting conditions and different disease severity levels, selecting a total of 2,931 images. The dataset was split into a training set and validation set at a ratio of 8:2. Labelimg software was used for labeling, where rectangular bounding boxes were added to each object, and a category label was assigned to each box. The label information included the category index or name of the object, the normalized coordinates of the bounding box center relative to the image dimensions, and the normalized width and height of the bounding box relative to the image. This dataset is referred to as "Appleleaf" (Apple Leaf Disease). The specific distribution is shown in Table I.

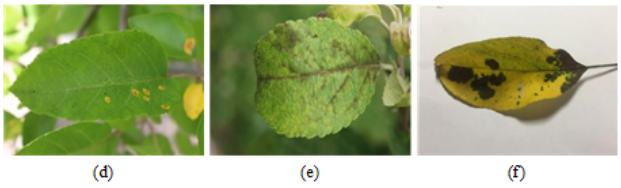

Fig. 2. ATLDSD apple leaf disease segmentation dataset (d) rust disease; (e) scab; And (f) brown spot

TABLE I
APPLELEAF IMAGE LEAF DISTRIBUTION

Type of disease	Quantity
Frogeye leaf spot	495
Powdery mildew	576
Scab disease	576
Brown spot	376
Health	432
Rust	496
Total	**2931**

B. Data augmentation

Considering the impact of complex orchard environments on data collection, it is crucial to account for scenarios where multiple environmental factors may interfere with disease detection [12]. By enhancing the dataset, we can simulate and expose the model to various conditions that may exist in apple orchards. This comprehensive data augmentation strategy improves the model's robustness and generalization ability, enabling it to better handle real-world complexities and improve the accuracy and reliability of disease detection.

We applied data augmentation techniques such as image rotation, scaling, and mosaic transformation, and simulated different weather conditions to create apple leaf datasets under low-light, overexposure, and fog conditions. Additionally, recognizing the possibility of multiple diseases affecting the same leaf, we used Photoshop for special image manipulations to reflect these conditions. Through this augmentation process, we expanded the dataset to 6,150 images, enhancing the model's ability to learn more features and improve generalization. The result enhanced images are shown in Fig.3.

C. Description of YOLOv8 algorithm

YOLOv8 offers a new state-of-the-art (SOTA)model designed to meet the needs of various scenarios and supports tasks such as image classification, object detection, instance segmentation, and pose detection [13]. YOLOv8 is a fast and accurate object detection algorithm that uses a single-stage detection approach and an anchor-based method for detecting objects. It employs a powerful backbone network

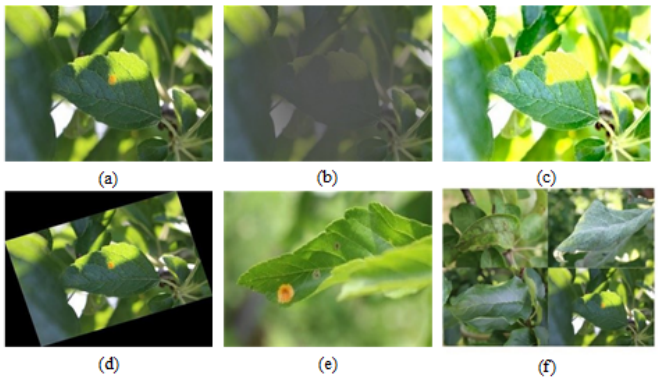

Fig. 3. Enhanced image (a) original image (b) fog (c) exposure (d) rotation (e) Mosaic (f) multi-disease

and a feature pyramid network to extract image features, along with an attention mechanism and an improved loss function. This enables the model to comprehensively consider factors like bounding box position, category prediction, and confidence, resulting in high performance and efficiency in object detection tasks.

Considering the real-time performance and accuracy requirements of the task, YOLOv8 has higher accuracy and faster speed in object detection while maintaining the characteristics of lightweight, so YOLOv8 is selected as the baseline model of this study.

The network structure of YOLOv8 is mainly composed of three parts: backbone network, neck network and detection head.

Backbone network: a series of convolution and deconvolution layers are used to extract features. At the same time, residual connection and bottleneck structure are also used to reduce the size of the network and improve performance. In this part, the C2f module is used as the basic constituent unit. Compared with the C3[14] module of YOLOv5, the C2f module has fewer parameters and better feature extraction ability.

Neck network: Multi-scale feature fusion technology is used to fuse the feature maps from different stages of Backbone to enhance the feature representation ability. Specifically, the Neck part of YOLOv8 includes a SPPF module, a PAA[15] module, and two PAN[16] modules.

Detection head: It is responsible for the final object detection and classification tasks, including a detection head and a classification head. The detection head contains a series of convolutional and deconvolution layers to generate detection results; The classification head uses global average pooling to classify each feature map Fig.4. shows the YOLOv8 network model.

III. THE PROPOSED ALGORITHM

A. Multi-Scale Progressive Feature Network

The proposed Multi-Scale Progressive Feature Network structure is illustrated in Fig.5. The first stage involves the fusion of two features with different resolutions. As the feature extraction process progresses from the bottom to the top of the backbone network, we gradually integrate high-level features related to leaf diseases. Through asymptotic fusion, we interactively combine the semantic information from low-level features with high-level features. To ensure

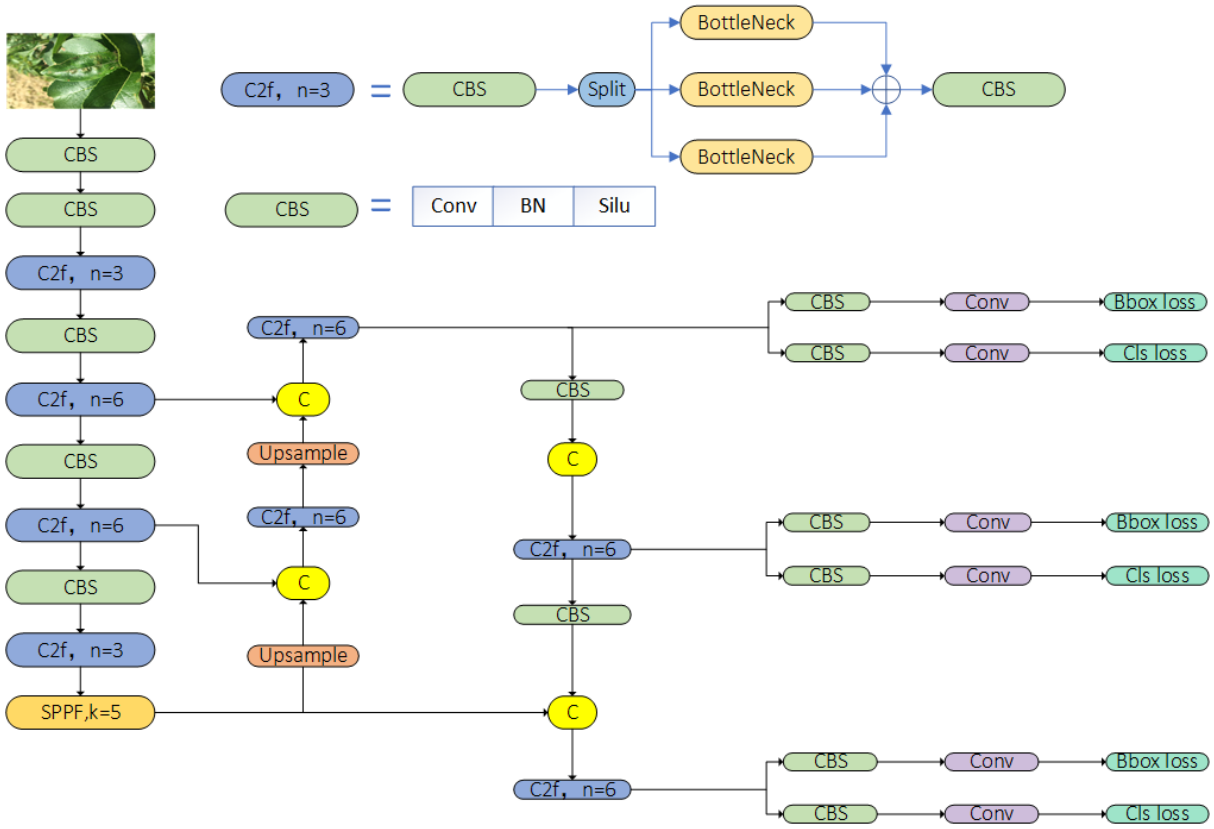

Fig. 4. YOLOv8 network model diagram

dimensional alignment and facilitate feature fusion, we utilize 1×1 convolutions and bilinear interpolation methods for upsampling the features. Additionally, we apply various convolution kernels and strides for downsampling, based on the required downsampling rate.

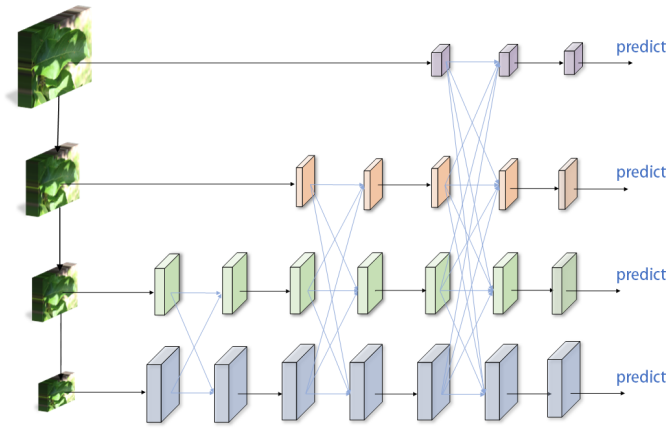

Fig. 5. Multi-Scale Progressive Feature Network (MSPN)

To optimize the multi-level feature fusion process, this study introduces a new data-driven method called Adaptive Spatial Feature Fusion (ASFF). ASFF incorporates learnable weights that dynamically adjust the contribution of feature maps from different scales. By learning these weights, the model can adaptively select and fuse features based on the disease characteristics at various scales. Additionally, ASFF employs a spatial alignment operation to ensure that feature maps of different scales maintain the same spatial resolution during fusion. Given the challenge of detecting small-sized lesions in apple leaf diseases, we added a small object detection layer to YOLOv8. This allows us to perform feature fusion across four levels of features with varying resolutions. We applied 2x2 convolution with a stride of 2 for 2× downsampling and 4x4 convolution with a stride of 4 for 4× downsampling, among others. After the feature fusion, we used four residual units to continue learning features, similar to ResNet, with each residual unit comprising two 3x3 convolutions. Recognizing that features from different levels contribute differently to disease characteristics, Adaptive Spatial Feature Fusion (ASFF) dynamically allocates spatial weights to enhance critical levels. By assigning varying degrees of importance to these features, ASFF improves the model's ability to capture and represent subtle and complex patterns of leaf diseases. This process not only boosts the recognition accuracy for key disease indicators but also ensures that less relevant features are downweighted, thereby optimizing the model's focus on significant areas, the Adaptive Spatial Feature Fusion structure demonstrated in Fig.6.

The figure shows the fusion process of three levels of features. Since the features of the three layers are fused, let the feature vectors representing the positions (i, j) from layer n to layer l be represented. The resulting feature vector, represented as y_{ij}^{l}, is obtained through adaptive spatial fusion of multi-level features, consisting of a linear combination of feature vectors $x_{ij}^{n \to l}$ is shown in equation (1).

$$y_{ij}^{l} = \alpha_{ij}^{l} \cdot x_{ij}^{1 \to l} + \beta_{ij}^{l} \cdot x_{ij}^{2 \to l} + \gamma_{ij}^{l} \cdot x_{ij}^{3 \to l} \tag{1}$$

In the formula, the sum of $\alpha_{ij}^{l} \beta_{ij}^{1 \to l} \gamma_{ij}^{l} \alpha_{ij}^{l} + \beta_{ij}^{l} + \gamma_{ij}^{l} = 1$

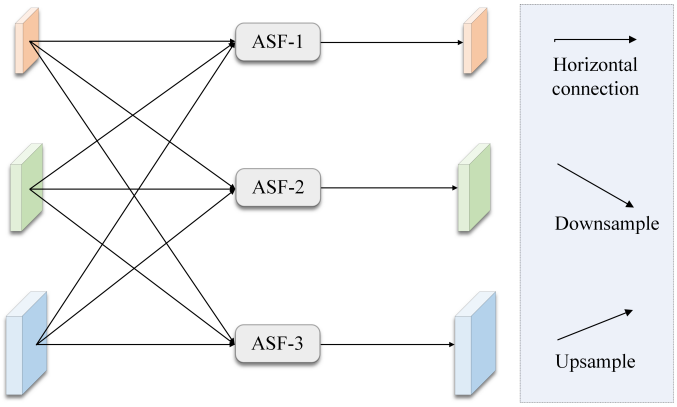

Fig. 6. Adaptive patial fusion(ASFF)

represents the spatial weight of the third level features at the first level. Under constraints, considering the difference in the number of fused features in each stage of MSPN, an adaptive spatial fusion module with a specific number of stages is implemented.

B. Global Self Attention Module

In apple leaf disease detection, different types of lesions can coexist in the same leaf, and some smaller areas of the lesion may be covered by surrounding healthy parts or other lesions, which often affects the accuracy of the detection results. To solve this problem, we choose to add an attention mechanism at the end of the backbone network, and the Global Self-attention module (GSAM) effectively extracts global information using sparse token region relationships. It can quickly focus on key areas and extract the features of the lesion when faced with complex environments. However, the non-overlapping space reduction used to reduce the number of tokens disrupts the spatial structure near the block boundaries and lowers the quality of the tokens. To address this issue, the Global Self-Attention Module (GSAM) introduces overlap spatial reduction (OSR) by using larger overlapping patches to represent the spatial structure near the patches better. Firstly, we perform a linear transformation operation on the input feature X, mapping it to the query vector Q. At the same time, we use the OSR module to downsample the input feature X. In this study, we reduce the spatial resolution of the feature map through depthwise separable convolution. The downsampled feature map s is convolved through a local convolution layer to generate a new feature map. Then, the convolved feature map is added to the original downsampled feature map s and mapped to the final key vector K and value vector V. This step is to preserve some of the original information while downsampling and enhancing the feature representation through residual connections. Finally, we reshape and transpose the QKV three vectors and send them to the Multi-Head Self Attention mechanism. As shown in Fig.7.

The OSR module is instantiated as a depthwise separable convolution, where stride follows PVT and the kernel size is equal to stride plus 3. It can also be expressed using equations (2) (3) (4) (5) :

$$Y = OSR(X) \tag{2}$$

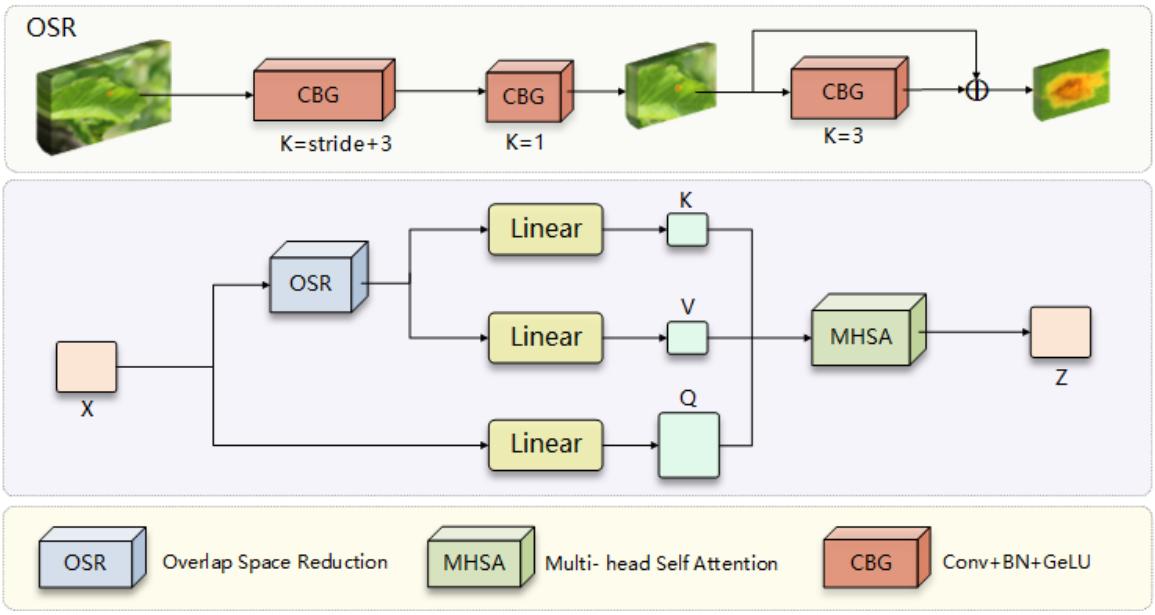

Fig. 7. Global Self-Attention Module (GSAM)

$$Q = Linear(X) \tag{3}$$

$$K, V = Split(Linear(Y + LR(Y))) \tag{4}$$

$$Z = Soft\max(\frac{QK^T}{\sqrt{d}} + B)V \tag{5}$$

Where $LR()$ represents a local refinement module instantiated by a 3x3 depth convolution, B is a position bias matrix, and d is the number of channels in each attention head. The global self-attention mechanism is added to capture the long-term distance information, and the sparse labeled region relationship is used to extract the global information efficiently, so as to improve the ability of the model to extract features.

C. Receptive Field-Focused Convolution Block

The spread and reproduction of pathogens, changes in environmental conditions, plant defense responses, fusion of disease spots, and the growth and aging of leaves themselves can all lead to changes in the shape and size of disease spots. To address this issue, we introduced Receptive Field-Focused Convolution Block (RFFconv) to replace the original convolutional blocks in the backbone network. RFFconv focuses on the spatial features of receptive fields. Firstly, adaptive average pooling is used to adjust the shape of the input feature map, and the Softmax function is introduced to calculate the weights w1, w2, and w3. By adding attention mechanism to adjust the parameters of the convolution kernel, it can dynamically respond to changes in different regions. Afterwards, three non shared convolutional layers, group1, group2, and group3, were generated for separate feature extraction and updating. Through the non shared features, different leaf diseases were learned. This flexibility enables the model to effectively capture and process the temporal changes of apple leaf lesions. By enhancing the understanding and processing of local areas, RFFconv can improve the accuracy and efficiency of lesion detection, and adapt to the dynamic changes in lesion morphology

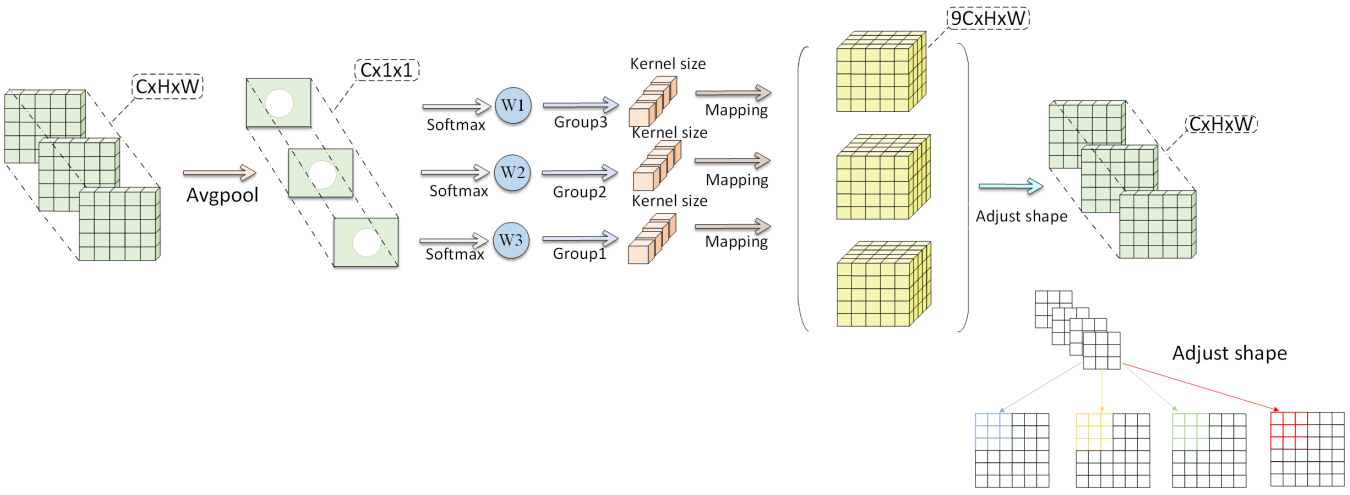

Fig. 8. Receptive Field-Focused Convolution Block (RFFconv)

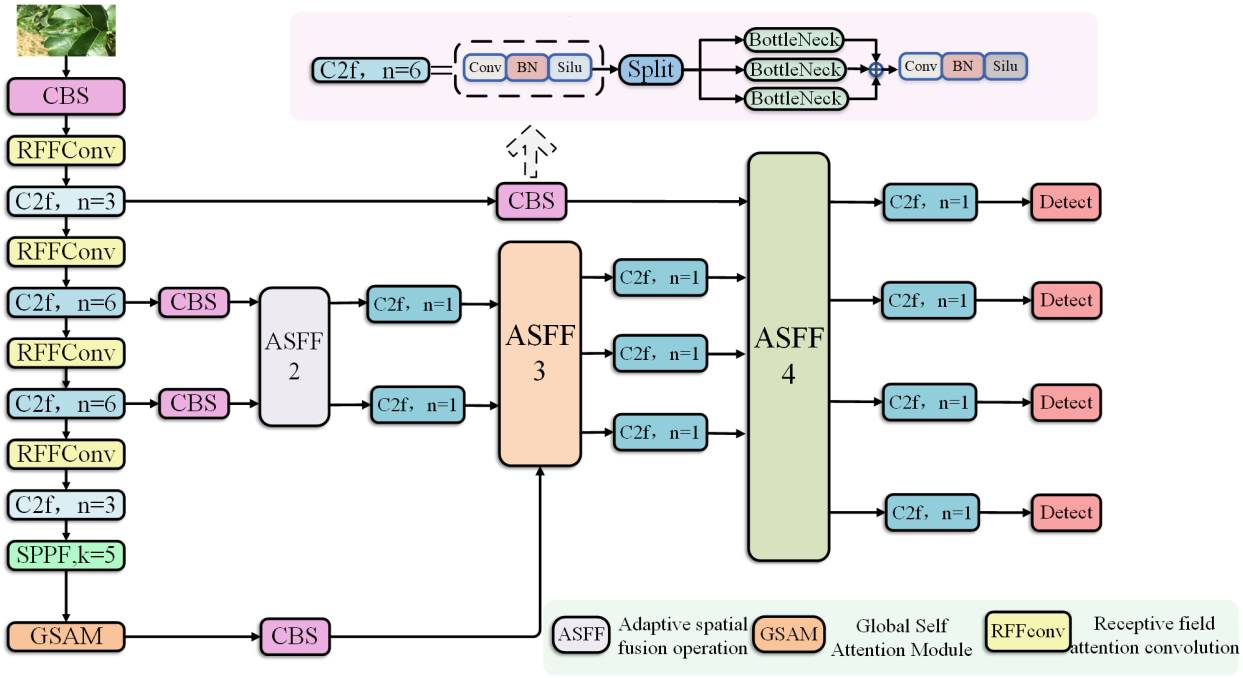

Fig. 9. MPGA-YOLOv8 model structure diagram

and size, thereby more accurately identifying and analyzing lesion features. The convolutional structure of receptive field attention demonstrated in Fig.8.

Receptive Field-Focused Convolution Block (RFFconv) is a method based on non-overlapping sliding Windows. When a 3×3 convolution kernel is used to extract features, each 3×3 window represents a receptive field slider. This method not only emphasizes the importance of different features within the receptive field slider, but also prioritises the spatial features of the receptive field. This method not only emphasizes the importance of different features, but also solves the problem of parameter sharing of traditional convolution kernels, thereby improving the learning ability of the model in complex image patterns. By dynamically generating the spatial features of the receptive field, RFFconv can more accurately capture and process the change characteristics of leaf disease spots, making the disease spot detection more accurate and efficient.

D. MPGA-YOLOv8 model

This study proposes MPGA-YOLOv8 for apple leaf disease detection, based on the YOLOv8n model. MPGA-YOLOv8 improves accuracy while maintaining detection speed. We enhance the Attention Mechanism (GSAM), Multi-Scale Progressive Feature Network (MSPN), and Receptive Field-Focused Convolution Block (RFFconv).We add the GSAM attention mechanism after the SPPF structure in the backbone network. This focuses on disease features, suppresses useless information, and improves detection accuracy. The neck network is replaced by MSPN, which fuses irregular disease features at multiple scales and combines context information to enhance the model's representation.RFFconv replaces the backbone network convolution. This adaptively adjusts the network's attention to objects of different scales, improving the detection of small and dense objects. Detecting small lesions early is challenging, so we add a small object detection layer [17] to YOLOv8n. This extracts small object features on high-resolution images, enhancing small object detection performance. MPGA-YOLOv8 is shown in Fig.9.

IV. EXPERIMENT

A. Experimental equipment and parameter Settings

The model runs on ubuntu system and uses pytorch deep learning framework for training and testing.Device Specifications: Intel(R) Xeon(R) Gold 6139M CPU @ 2.30GHz processor, 32GB RAM,NVIDIA GeForce RTX 3060 graphics card, 12GB video memory, CUDA version 11.6, cudnn version 8.9.5, python version 3.9. The image size was normalized to 640x640, the initial learning rate was set to 0.01, the learning rate was reduced by cosine annealing method, the number of training rounds epoch was set to 200, and the image batch size was 32. As shown in Table II.

TABLE II
HYPERPARAMETER SETTINGS

Hyperparameter Settings	value
imgsz	640x640
Workers	8
Lr0	Auto
Momentum	0.01
Epochs	200
Batch size	32
Patience	**50**

B. Evaluation Metrics

YOLOv8n model algorithm performance evaluation indicators include model accuracy, recall, mean Average Precision (mAP)[18], model size, floating-point arithmetic (FLOPS), FPS, etc., which are used to measure the accuracy and real-time performance of the model in object detection tasks. As the most common evaluation index, Precision represents the meaning that precision measures the accuracy

of the model in the samples predicted as positive samples. More attention is paid to the accuracy of the model predicted as positive samples, and the calculation formula (6) is as follows:

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

Recall is a key indicator of object detection, meaning that recall measures the proportion of positive samples correctly detected by the model, and more attention is paid to the coverage of positive examples by the model. The calculation formula (7) is as follows:

$$\text{Re}call = \frac{TP}{TP + FN} \tag{7}$$

Class Mean Average Precision (mAP) is the overall performance of the model in multi-classification tasks and object detection tasks, which is determined by precision and Recall. AP is the integral of precision and recall, and mAP is the average of AP. TP stands for the number of positive samples correctly identified by the model, FP stands for the number of negative samples incorrectly identified by the model as positive samples, and FN stands for the number of positive samples incorrectly identified by the model as negative samples. n is the number of disease species. Formula (8) and (9) are calculated as follows:

$$AP = \int_0^1 P(R)dR \tag{8}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \tag{9}$$

The number of floating-point operations required to process an image allows a fair comparison of the detection speed between different algorithms. FLOPs and FPS are used as evaluation metrics. For convolutional layers, the formula is as follows:

$$FLOPS = 2HW(C_{in}k^2 + 1)C_{out} \tag{10}$$

Where is the number of channels of the input tensor of the convolution layer, is the number of channels of the output tensor of the convolution layer, and K refers to the size of the convolution kernel $C_{\text{in}}/C_{\text{out}}$.

C. Display of results

Through the improvement of YOLOv8n model for apple leaf disease detection task, AppleLeaf dataset was used for training and testing, and the baseline model was compared. The class average precision is increased by 9.5% when iou=0.5, and it can be seen that brown spot, powdery mildew and scab diseases have a relatively large improvement, increasing by 15.9%, 10.7% and 7.8% respectively. At the same time, among all disease types, the recognition accuracy of scab disease is low, which has been greatly improved after improvement. The results are shown in Table III.

TABLE III
MODEL EVALUATION

Models	YOLOv8n(AP%)	MPGA-YOLOv8(AP%)
all	0.865	0.926
Brown spot	0.797	0.956
Frogeye leaf spot	0.957	0.935
Health	0.905	0.958
Powdery mildew	0.839	0.946
Rust	0.975	0.966
Scab	0.717	0.795

The improved model was tested under different natural backgrounds and compared with the baseline model, which was divided into haze weather, strong light conditions, normal weather and the presence of multiple diseases on leaves. The improved model MPGA-YOLOv8 accurately detected disease spots in these extreme environments, and the improved model paid more attention to small-size disease spots. The improved model can learn the disease characteristics well and adapt to the size changes of different diseases. The undetected disease spots are marked with red or yellow circles, and the baseline model of YOLOv8n is compared with the improved model, the results are shown in Fig.10.

The baseline model YOLOv8n and the improved model MPGA -YOLOv8 were trained for 150epochs respectively, and the accuracy (Presicion), Recall, class average precision mAP@0.5 and mAP@0.95 were compared, and the results are shown in Fig.11.

To further explore the interpretability of the model, this paper uses the Grad-CAM method to visualize the class activation maps, as shown in Fig 12. This approach allows us to visually observe the model's attention distribution during object detection and the degree of focus on different regions.In the experiment, we compared the YOLOv5s and YOLOv8n models, using the SPPF layer as the detection layer. We visualized and analyzed four representative leaf

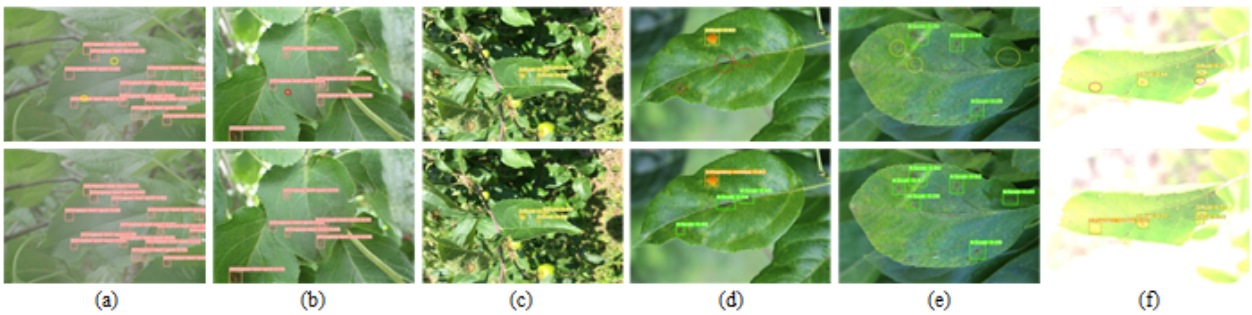

Fig. 10. Comparison of test charts before and after model improvement. (a) fog condition (b) frog eye leaf spot under normal weather condition (c) same plant multiple disease condition (d) rust disease under normal weather condition (e) scab disease under normal weather condition(f) strong light condition

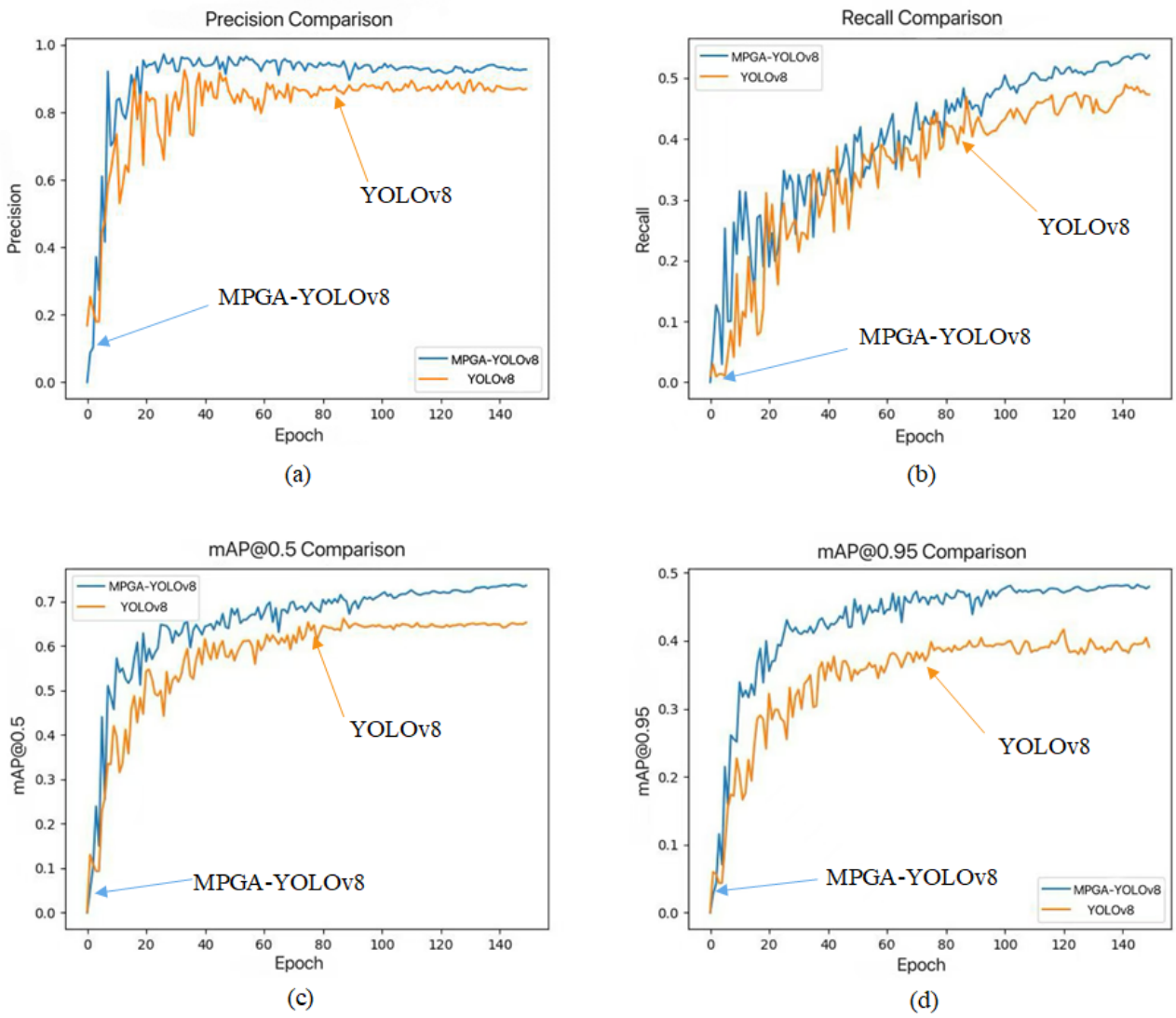

Fig. 11. Model Performance comparison plot (Precision, Recall, mAP@0.5, mAP@0.95)

disease images. The comparison of Grad-CAM heatmaps clearly shows that the MPGA-YOLOv8 model exhibits more concentrated and relevant activations in key task areas, accurately identifying and locating disease regions. In contrast, the other models have more scattered attention points and lower accuracy, indicating that the MPGA-YOLOv8 model performs better in task understanding and detection.

D. Ablation experiment

To further validate the effectiveness of the experiment and improve the performance of the algorithm, we conducted ablation experiments by comparing the individual modules and controlling variables to maintain input consistency. All images and training epochs remained unchanged. To better evaluate the experimental results of our model, we used mean Average Precision (mAP@0.5) and Recall as our performance metrics, which are critical for accurate disease identification and localization. In the experiment, we sequentially added Multi-Scale Progressive Feature Networks (MSPN), Receptive Field Focusing Convolution Blocks (RFFconv), Global Self-Attention Modules (GSAM), and Small Target Detection Layers (STD layer) to the baseline model YOLOv8n. The experimental results showed that, without any algorithmic improvements, the baseline model achieved an mAP of 64.8% and a Recall of 40.5%. After adding the Multi-Scale Progressive Feature Network (MSPN), the mAP increased by 4.4%,and Recall increased by 0.7%. This significant improvement suggests that the fusion of multi-level features effectively enhanced the lesion characteristics, though the localization performance for the disease was not significantly improved.Next, we sequentially added the Receptive Field Focusing Convolution Block (RFFconv) and the Global Self-Attention Module (GSAM). The results showed that when all three modules were updated together, the model achieved the best performance, with the mAP increasing by 7.2% and Recall increasing by 7.1%. This indicates that RFFconv enhanced the flow of information between different layers of the deep network, expanding the receptive field while focusing more on the changes in lesion features, while GSAM helped the model better understand the global context of the image, leading to more accurate object recognition and localization.Finally, to improve the detection ability for small lesions, we incorporated the Small Target Detection Layer (STD layer), which ultimately increased the mAP

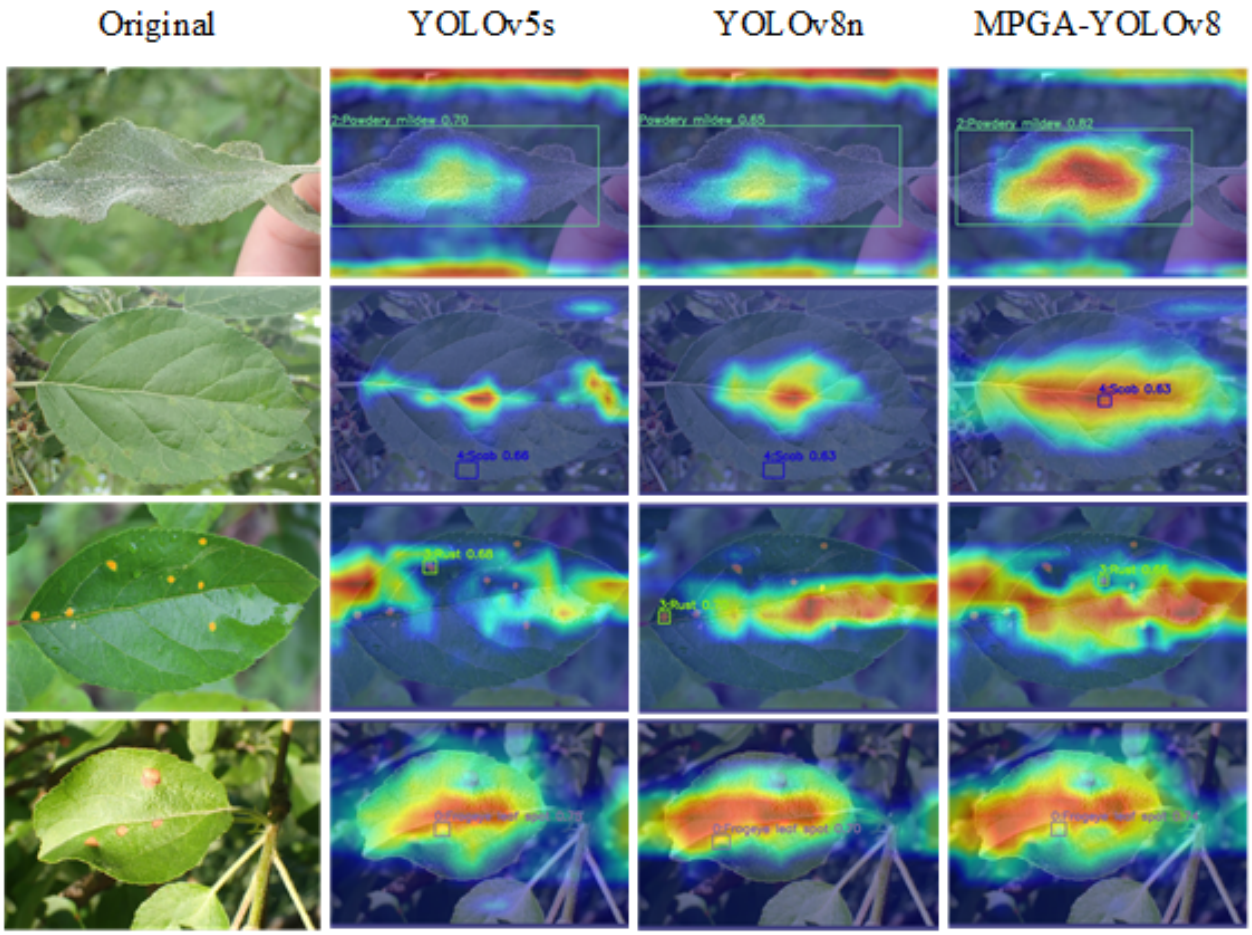

Fig. 12. Grad-CAM visualization results of different algorithm

by 9.5% and Recall by 12%. The model achieved its best performance. Therefore, the experimental results demonstrate that our proposed improvements are effective and applicable. The results are shown in Table IV.

TABLE IV
ABLATION EXPERIMENT

MSPN	RFFconv	GSAM	STD layer	mAP@0.5	Recall
				64.8%	40.5%
✓				69.2%	41.2%
	✓			67.3%	42.9%
		✓		68.1%	42.6%
✓	✓			66.3%	42.7%
✓		✓		67.4%	43.2%
✓	✓	✓		72.0%	47.6%
✓	✓	✓	✓	74.3%	52.5%

E. Comparison of state-of-the-art models

This study chooses SSD, Faster RCNN, YOLOv3-tiny, Retinanet, EfficientDet, YOLOv5, YOLOv6, YOLOv8n and other YOLO series detection models for comparison. The model accuracy mAP@0.5, model size and the number of images processed by the model per second are used as the evaluation criteria. by comparing the experimental results of different models, MPGA-YOLOv8 achieves the best detection accuracy with an mAP@0.5 of 0.743. This result is significantly higher than other models. Its model size is only 6.07 MB, and the computation is 16.1 GFLOPS, showing efficient performance. In comparison, SSD and Faster RCNN have mAPs of 0.534 and 0.359, with larger model sizes and higher computation requirements. They are less efficient than MPGA-YOLOv8. Lightweight models like YOLOv5 and YOLOv8n have advantages in size and computation but fall behind MPGA-YOLOv8 in accuracy. MPGA-YOLOv8 balances accuracy, model size, and computation, making it the most effective model.The experimental results are shown in Table V. Furthermore, considering the practical needs of disease detection tasks, MPGA-YOLOv8 achieves superior inference speed and higher FPS performance. The comparison with these advanced models further demonstrates the superiority of the MPGA-YOLOv8 model, whose lower model complexity makes it more suitable for deployment on edge devices. The result of inference speed is shown in Fig.13.

TABLE V
MODEL COMPARISON

Models	mAP@0.5	Model size (MB)	FLOPS(G)
SSD	0.534	32.07	123.3
Faster RCNN	0.359	37.66	176.04
YOLOv3-tiny	0.645	23.24	19.0
Retinanet	0.625	26.15	3.7
EfficientDet	0.601	7.16	9.5
YOLOv5	0.633	4.78	7.2
YOLOv6	0.600	8.08	11.9
YOLOv8n	0.648	5.76	8.2
MPGA-YOLOv8	0.743	6.07	16.1

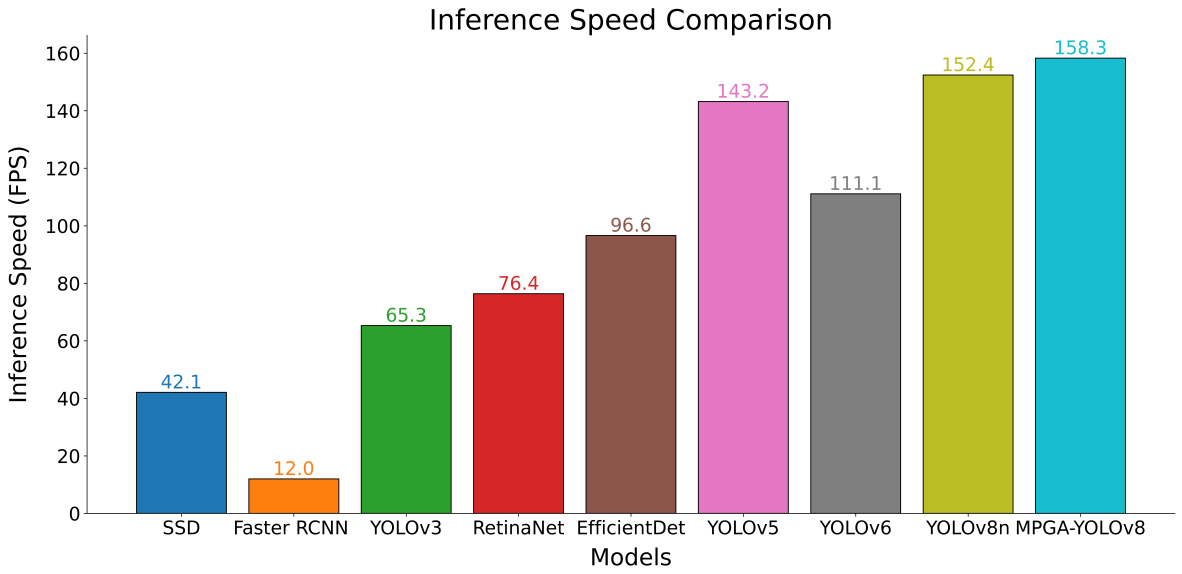

Fig. 13. Inference Speed Comparison Chart

V. Conclusion

Due to the specific challenges of apple leaf disease detection, the small size of frogeye leaf spots and rust spots complicates the task[19]. Small lesions are difficult for models to accurately identify and require extremely high resolution and fine feature extraction capabilities to ensure precise localization and detection of the diseases. Additionally, the irregular shape of scab spots and the evolving symptoms of the disease over time contribute to this complexity. Diseases at different developmental stages may exhibit varied visual characteristics, further complicating model training and detection.to tackle these challenges, this study proposes the MPGA-YOLOv8 object detection model, based on an improved YOLOv8n. We combine the original neck network with the Multi-Scale Progressive Feature Network (MSPN), enhancing the model's performance in detecting objects of varying sizes through feature fusion and enhancement techniques. The Global Self Attention Module (GSAM) has been incorporated at the end of the backbone network to improve the model's ability to represent disease points by capturing global information and long-range dependencies. Using the Receptive Field-Focused Convolution Block (RFFconv) in place of inner convolutions in the backbone allows for more accurate capture and processing of the time-varying characteristics of leaf lesions. Finally, to address the challenge of detecting small target lesions, we added a small target detection layer to further extract features of these lesions. Through these improvements and optimizations, we successfully addressed the detection of six leaf disease samples (including healthy samples), focusing on issues related to varying lesion sizes, the detection of small target lesions, complex environments (fog, strong light, weak light), and missed or false detections caused by multiple diseases present on the same plant[20].

References

[1] B. E. Juniper, R. Watkins, and S. A. Harris, "The origin of the apple," in *Eucarpia Symposium on Fruit Breeding and Genetics 484*, 1996, pp. 27–34.

[2] H.-D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.

[3] J. Heaton, "An empirical analysis of feature engineering for predictive modeling," in *SoutheastCon 2016*. IEEE, 2016, pp. 1–6.

[4] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of deep learning-based object detection," *IEEE access*, vol. 7, pp. 128 837–128 868, 2019.

[5] Y. Li, S. Sun, C. Zhang, G. Yang, and Q. Ye, "One-stage disease detection method for maize leaf based on multi-scale feature fusion," *Applied Sciences*, vol. 12, no. 16, p. 7960, 2022.

[6] X. Chen, X. Ye, M. Li, Y. Lou, H. Li, Z. Ma, and F. Liu, "Cucumber leaf diseases detection based on an improved faster rcnn," in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, vol. 6. IEEE, 2022, pp. 1025–1031.

[7] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *Ieee Access*, vol. 7, pp. 59 069–59 080, 2019.

[8] G. LIU, G. HU, G. L. B. H. ER, T. ZHAO, Y. DONG *et al.*, "Detection of grape leaf diseases and insect pests based on improved yolov3," *Microelectronics & Computer*, vol. 40, no. 2, pp. 110–119, 2023.

[9] L. Sun, G. Hu, C. Chen, H. Cai, C. Li, S. Zhang, and J. Chen, "Lightweight apple detection in complex orchards using yolov5-pre," *Horticulturae*, vol. 8, no. 12, p. 1169, 2022.

[10] S. Liu, Y. Qiao, J. Li, H. Zhang, M. Zhang, and M. Wang, "An improved lightweight network for real-time detection of apple leaf diseases in natural scenes," *Agronomy*, vol. 12, no. 10, p. 2363, 2022.

[11] Q. Yang, S. Duan, and L. Wang, "Efficient identification of apple leaf diseases in the wild using convolutional neural networks," *Agronomy*, vol. 12, no. 11, p. 2784, 2022.

[12] P. Enkvetchakul and O. Surinta, "Effective data augmentation and training techniques for improving deep learning in plant leaf disease recognition," *Applied Science and Engineering Progress*, vol. 15, no. 3, pp. 3810–3810, 2022.

[13] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[14] H. Park, Y. Yoo, G. Seo, D. Han, S. Yun, and N. Kwak, "C3: Concentrated-comprehensive convolution and its application to semantic segmentation," *arXiv preprint arXiv:1812.04920*, 2018.

[15] E. Chai, L. Ta, Z. Ma, and M. Zhi, "Erf-yolo: A yolo algorithm compatible with fewer parameters and higher accuracy," *Image and Vision Computing*, vol. 116, p. 104317, 2021.

[16] Y. Du, N. Pan, Z. Xu, F. Deng, Y. Shen, and H. Kang, "Pavement distress detection and classification based on yolo network," *International Journal of Pavement Engineering*, vol. 22, no. 13, pp. 1659–1672, 2021.

[17] A. Benjumea, I. Teeti, F. Cuzzolin, and A. Bradley, "Yolo-z: Improving small object detection in yolov5 for autonomous vehicles," *arXiv preprint arXiv:2112.11798*, 2021.

[18] K. He, Y. Lu, and S. Sclaroff, "Local descriptors optimized for average precision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 596–605.

[19] M. Sebastian, M. Suchithra, and C. M. Antony, "Apple leaf disease detection: Machine learning & deep learning techniques," in *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*. IEEE, 2023, pp. 1–5.

[20] J. Li, X. Zhu, R. Jia, B. Liu, and C. Yu, "Apple-yolo: A novel mobile terminal detector based on yolov5 for early apple leaf diseases," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2022, pp. 352–361.