PAME-YOLO: A MODEL FOR APPLE LEAF LESION DETECTION IN COMPLEX ENVIRONMENTS BASED ON IMPROVED YOLOv8s

/

PAME-YOLO：一种适用于复杂环境的基于改进 YOLOv8s 的苹果叶片病斑检测模型

Yuansheng BING[1)], Xiao YU[*,1,2)], Zeqi LIN[1)], Feng YU[1)]

[1)] School of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong/China

[2)] Institute of Modern Agricultural Equipment, Shandong University of Technology, Zibo, Shandong/China

Tel: +8613163680678; E-mail: neaufish@sdut.edu.cn

Corresponding author: Xiao Yu

DOI: https://doi.org/10.35633/inmateh-75-97

Keywords: *YOLOv8s，apple leaf lesions, target detection, attention mechanism, complex environments*

ABSTRACT

The detection of apple leaf lesions in complex environments is hindered by several factors, such as the small size of lesion areas, variability in lighting conditions, and occlusions caused by overlapping leaves. These issues significantly limit the performance of existing detection models. Therefore, an enhanced detection algorithm for apple leaf lesions, termed PAME-YOLO, is proposed in this study, building upon the YOLOv8s framework. First, the main convolutional module is reconstructed using the Parallelized Patch-Aware Attention Module (PPA) while fusing Efficient Multi-Scale Attention (EMA). This effectively strengthens the model's capacity to localize small target lesions in complicated environments. Second, an Attention-based Intra-scale Feature Interaction (AIFI) is introduced into the feature extraction network to replace the Spatial Pyramid Pooling-Fast (SPPF) module, which better captures the subtle features of apple leaf lesions. Next, the downsampling enhancement module is designed to mitigate information loss during the original downsampling process, which contributes to a significant improvement in detection precision. Finally, the Efficient Head is designed, a lightweight and efficient detection head that lowers parameter count and computational intricacy without sacrificing accuracy. Compared with YOLOv8s, the proposed model delivered a notable enhancement in performance, with precision (P) increasing by 0.8 points and recall (R) by 1.5 points. The mAP@0.5 achieved 91.4%, which is 1.5 percentage points higher than that of YOLOv8s. Meanwhile, the mAP@0.5:0.95 rose to 56.4%, reflecting an increase of 1.4 percentage points. The improved model realizes the accurate detection of apple leaves lesions in complicated surroundings, offering reliable technical assistance for disease prevention and contributing to the development of the apple industry.

摘要

在复杂环境中，苹果叶片病斑的检测受到诸多因素的影响，如病斑区域尺寸较小、光照条件变化以及叶片重叠造成的遮挡等问题。这些因素严重限制了现有检测模型的性能。为此，本文在 YOLOv8s 算法的基础上提出了一种称为 PAME-YOLO 的苹果叶片病斑检测算法。首先，本文使用并行补丁感知注意力模块同时结合高效多尺度注意力机制对主干卷积模块进行重构，这有效增强了模型在复杂环境中对小目标病斑的定位能力。其次，本文在特征提取网络中引入基于注意力的尺度内特征交互模块，来替换原有的快速空间金字塔池化模块，以更好地捕捉苹果叶片病斑的细微特征。随后，本文设计了新的下采样增强模块，以弥补原有下采样过程中的信息丢失，从而显著提高检测精度。最后，我们设计了一种轻量高效的检测头 Efficient Head，该检测头能够在保持精度的同时降低模型参数和计算复杂度。与 YOLOv8s 相比，所提出的模型在性能上取得了显著提升，精确率（P）提高了 0.8 个百分点，召回率（R）提高了 1.5 个百分点，mAP@0.5 达到了 91.4%，比 YOLOv8s 高出了 1.5 个百分点，同时 mAP@0.5:0.95 达到了 56.4%，提高了 1.4 个百分点。综上所述，改进后的模型实现了在复杂环境下对苹果叶片病斑的精准检测，为病害防控提供了可靠的技术支持，助力了苹果产业的可持续发展。

INTRODUCTION

Apple is a major cash crop that is widely consumed across the globe *(Bai et al., 2021)*. However, it is susceptible to diseases, particularly those affecting the leaves. Leaf infections can significantly disrupt physiological metabolism and photosynthesis, directly impairing apple growth and harvest. Consequently, prompt and precise identification of apple leaf diseases is essential for effective orchard management. This not only enables growers to prevent the spread of diseases and improve fruit quality and yield but also contributes to substantial economic and environmental benefits.

In earlier years, manual observation was the primary method for identifying apple diseases in most orchards and farms. However, this approach is time-consuming, prone to misdiagnosis, and increasingly inadequate for meeting the demand for fast and accurate disease identification. As artificial intelligence continues to evolve, object detection techniques powered by deep learning *(Yann et al., 2015)* have found increasing application in the agricultural situations. According to the processing pipeline, models for object detection are typically categorized into two distinct groups: two-stage and one-stage detectors. The most representative example of a two-stage detection method is R-CNN *(GIRSHICK et al., 2014)*. *Gong et al., (2023)* suggested an improved Faster R-CNN algorithm. This algorithm has an average accuracy of 63.1% on an annotated apple leaf disease dataset, which surpasses other target detecting techniques. *Zhang et al., (2021)* presented a soybean leaf disease detection model named MF3R-CNN, which employs skip connections between multiple layers in the feature extraction network to facilitate multi-feature fusion, thereby effectively fulfilling the necessities of object detection tasks. Despite the great detection accuracy provided by two-stage algorithms, their training and inference processes are time-consuming, which limits their suitability for deployment in intelligent agricultural equipment.

Compared to two-stage detectors, one-stage detection methods are more appropriate for practical applications due to their better scalability and faster inference speed. Among the most well-known one-stage detection methods is the YOLO series, which has gained widespread deployment in the agriculture industry. *Abulizi et al., (2024)* integrated lightweight dynamic Sampling (DySample) to enhance small lesion feature extraction and employed Margin Penalty Distance Intersection over Union (MPDIoU) for precise localization of overlapping lesion boundaries. These enhancements achieved higher accuracy in tomato leaf disease recognition. To identify apple leaf diseases, *Li et al., (2023)* proposed an improved YOLOv5s model by introducing a Bi-Directional Feature Pyramid Network (BiFPN), a Transformer module, and the Convolutional Block Attention Module (CBAM) to reduce background interference. The model achieved an average detection accuracy of 84.3% in natural environments. *Gao et al., (2024)* replaced the traditional convolution and C2f structure with GhostConv and C3Ghost, respectively. They also incorporated the Global Attention Mechanism (GAM) and a BiFPN to enhance the detection of small apple leaf lesions in complicated environments.

Although the studies described above have made some progress, several problems in detecting apple leaf diseases remain unresolved. First, some disease spots are small, and different diseases may exhibit similar features, resulting in the model struggling to detect the lesions accurately. Additionally, in real-world cultivation environments, factors such as leaf occlusion and uneven lighting can reduce the model's capability to concentrate on diseased areas, limiting detection performance. Moreover, some studies have placed excessive emphasis on improving precision, while overlooking the increased model intricacy and computing expenses.

To overcome these issues, PAME-YOLO was developed, an enhanced YOLOv8s-based model for detecting apple leaf disease spots. The following contributions were made by this paper:

1. The C2f-PE module is designed by incorporating the PPA *(Xu et al., 2024)* module and the EMA *(Ouyang et al., 2023)* attention mechanism to further enhance the capability of identifying small lesions and differentiating similar features.

2. The SPPF module is substituted with the scale interaction module AIFI *(Zhao et al., 2024)*, which strengthens the high-level feature extraction, improves detection performance, and reduces redundant computation.

3. The downsampling enhancement module MPC is designed to improve the model's attention to small lesions in complex backgrounds, enabling it to better preserve key contextual information and enhance detection accuracy.

4. Aiming at the problem of high computational and large parametric quantities of the original detection head, the Partial Convolution (PConv) *(Chen et al., 2023)* was employed to design the Efficient Head detection head, which increases detection efficiency while lowering computing costs and model complexity.

MATERIALS AND METHODS

Dataset

The raw images of the apple leaf disease in this research are acquired by the publicly available dataset AppleLeaf9 *(Yang et al., 2022)*. Approximately 94% of the images in this dataset were taken in field settings, which ensures that the collected image data meet the requirements for complex backgrounds. From the dataset, 1,607 images of three common apple leaf spot diseases—Alternaria leaf spot, rust, and grey spot—were selected as the original image data. To address the limited original dataset, the data augmentation was applied to increase the images to 8952. After that, the augmented images were separated into training, validation, and test sets at a proportion of 8:1:1, with disease locations and categories annotated using the Labelimg annotation tool.

YOLOv8

YOLOv8 is a strong visual recognition framework made to handle a range of computer vision tasks, including object detection, image classification, and instance segmentation *(Wang et al., 2025)*. Compared to its predecessors YOLOv5 and YOLOv7, YOLOv8 delivers improved recognition precision along with accelerated inference performance. Its architecture comprises three primary components: the backbone, the neck, and the detection head *(Tian et al., 2024)*. In feature extraction and fusion stages, YOLOv8 substitutes the C2f structure for the C3 module that was utilized in YOLOv5, facilitating richer gradient flow and improving feature representation. The classification and detection tasks are separated in the head by YOLOv8's decoupled structure, enhancing detection performance. Furthermore, it replaces the anchor-based mechanism of YOLOv5 with an anchor-free approach, giving the model greater flexibility and efficiency in identifying objects of different sizes and forms.

Improved YOLOv8 Algorithm

To address challenges such as the small size of disease spot features, high similarity among different lesions, reduced detection accuracy in complicated cultivation environments, and the extensive parameters, this study suggests an improved model based on YOLOv8s, named PAME-YOLO, as illustrated in Fig. 1.

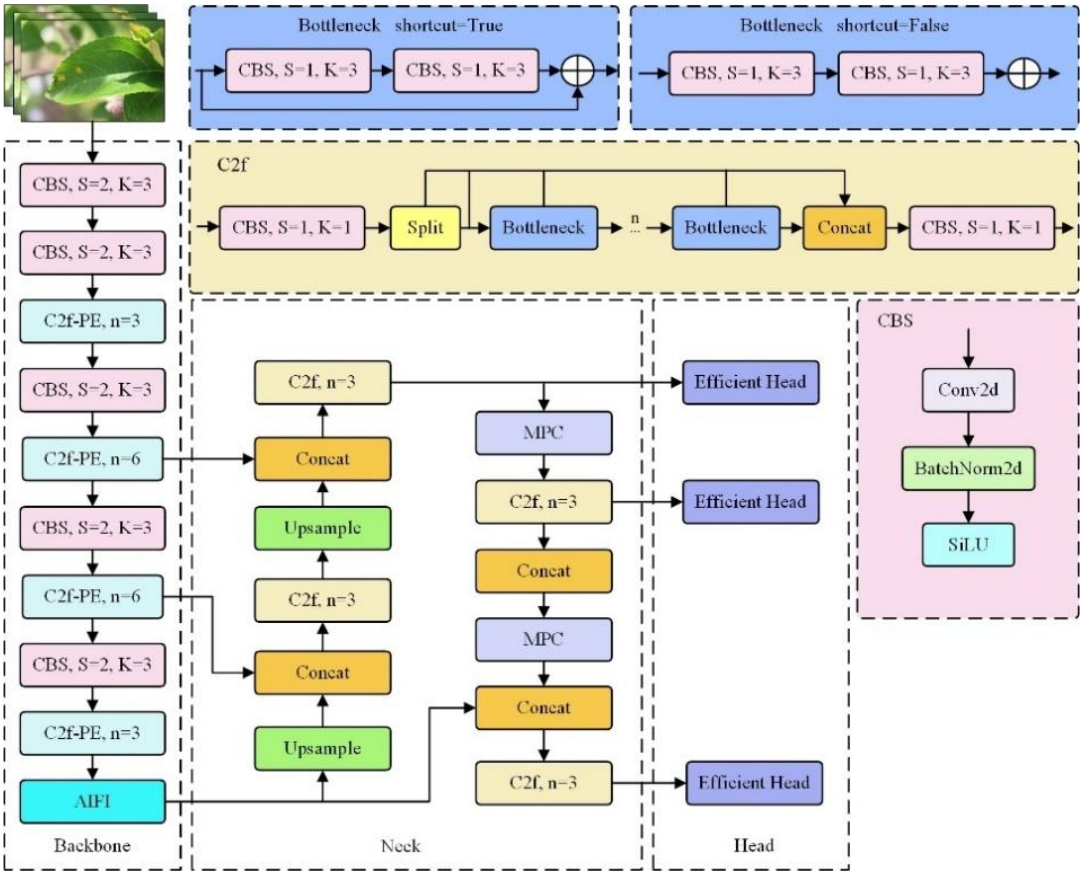

Fig. 1 - Overall architecture of the PAME-YOLO

C2f-PE

The C2f module can enable multi-scale feature extraction and fusion. Its multi-branch design strategy enhances the network's adaptability and representation capability. However, its ability to detect small targets and distinguish similar features remains limited. Therefore, this study redesigns and enhances the C2f module by integrating the PPA module and the EMA attention mechanism. The enhanced C2f-PE module is displayed in Fig. 2.

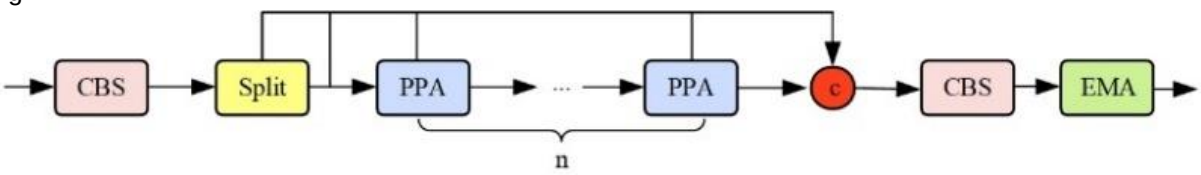

Fig. 2 - Structure of C2f-PE Module

Multiple downsampling operations can lead to the loss of details about small targets and missed detections. To make the model better at locating small disease spots, this paper incorporates the PPA module from the context fusion network HCF-Net into the C2f module. As illustrated in Fig. 3, the PPA module is made up of two essential parts: a multi-branch feature extraction architecture and an attention mechanism. The main benefit of PPA is its multi-branch feature extraction strategy. This method effectively increases the accuracy of detecting tiny disease spots on apple leaves by employing parallel branches, which extract features at different sizes and levels. The initial step in the feature extraction procedure is to convert the input tensor $F \in \mathbb{R}^{H' \times W' \times C}$ into $F' \in \mathbb{R}^{H' \times W' \times C'}$ using point-wise convolution. Then, F' is handled through three distinct parallel paths, which respectively generate the local feature tensor $F_{local} \in \mathbb{R}^{H' \times W' \times C'}$, the global feature tensor $F_{glocal} \in \mathbb{R}^{H' \times W' \times C'}$, and the linear feature tensor $F_{conv} \in \mathbb{R}^{H' \times W' \times C'}$. Lastly, the fused feature map $\tilde{F} \in R^{H' \times W' \times C'}$ is obtained by adding the three tensors. After multi-branch feature extraction, an attention module is applied to produce the final output. The attention module is constituted by a sequence of channel attention *(Wang et al., 2020)* and spatial attention mechanisms *(Woo et al., 2018)*. This design is particularly effective for detecting small disease spots on apple leaves and suppressing background noise, leading to improved accuracy and robustness of the model. The parameter p, which defines the patch size, serves to differentiate local and global branches, thereby promoting spatial feature fusion and displacement encoding *(Bi et al., 2025)*.

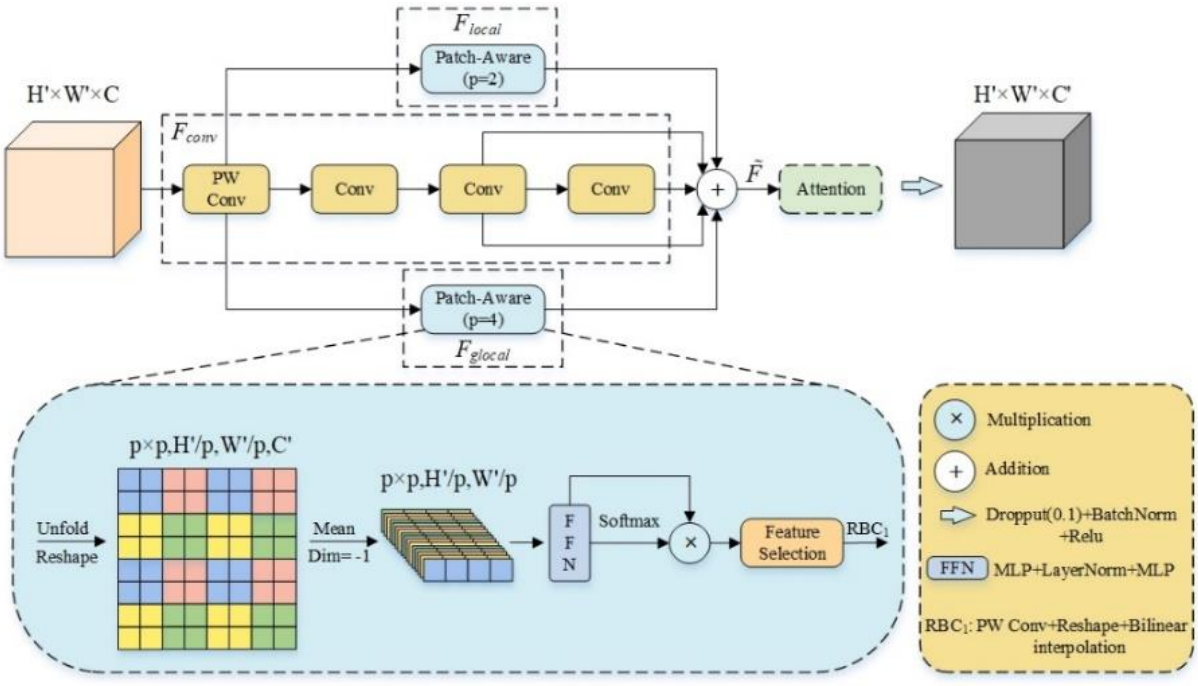

Fig. 3 - Parallelized Patch-Aware Attention Module

A certain degree of feature similarity exists both among different types of disease spots and between disease spots and surrounding objects, which can result in false detections. This issue is related to the limitation in the model feature extraction capability or the insufficient ability to select extracted feature information. To strengthen the model's feature extraction capacity, this paper also incorporates the EMA attention mechanism after the C2f module.

Fig. 4 presents the EMA module, which adopts cross-spatial learning to achieve efficient scale-aware attention. It reshapes part of the channels into the batch dimension and applies grouping in the channel dimension without requiring dimensionality reduction. This successfully stops channel feature information from being lost while reducing computational cost, and it features high accuracy and a low parameter count.

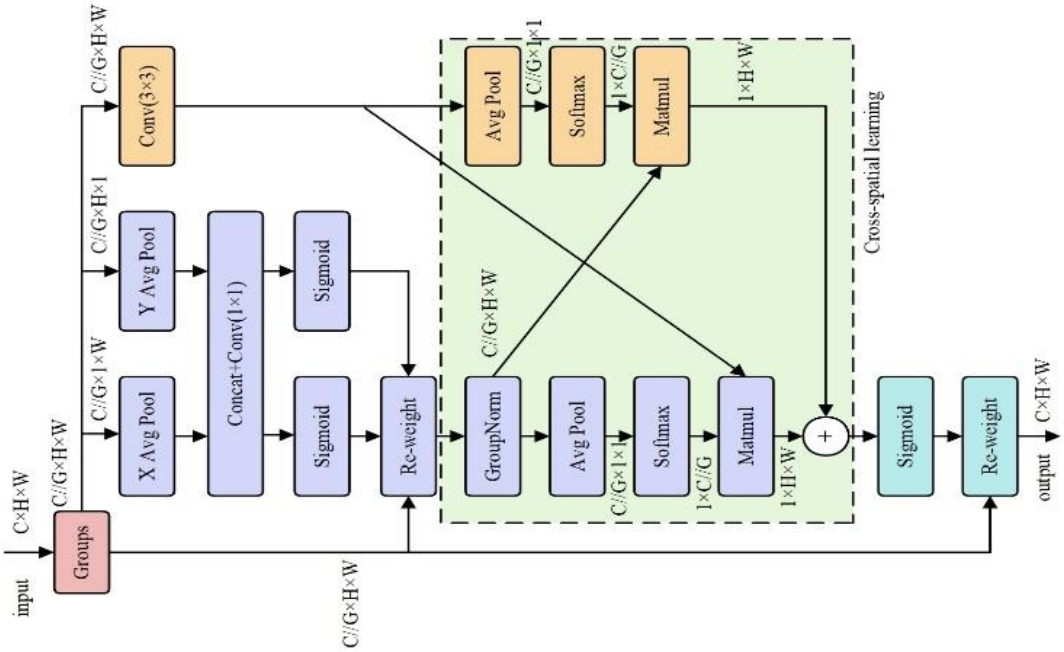

Fig. 4 - EMA attention mechanism framework diagram

AIFI

The SPPF module is an essential component of YOLOv8, enabling multi-scale feature fusion to enhance contextual information capture. However, it incurs a high cost of calculation. To mitigate this problem, the SPPF module is substituted with the AIFI module, which focuses on processing advanced image features. This contributes to enhanced detection accuracy while lowering computational cost. Compared to traditional multi-scale feature fusion methods, AIFI employs a single-scale Transformer encoder to focus feature fusion within the same scale. This helps capture finer-grained information and reduces the computational cost. Advanced features contain richer semantic content compared to low-level features, which have limited contribution due to insufficient semantic representation. As a result, the intra-scale interactions of lower-level features are redundant.

As seen in Fig. 5, the AIFI firstly linearizes the input 2D picture S5, converting it into a one-dimensional vector by arranging the rows sequentially. Then, a multi-head attention mechanism is employed, enabling the model to gather information from different spatial locations in the sequence, which strengthens its capacity to model long-range dependencies in the feature representation. The processed sequence is then combined with the original input for layer normalization. Afterward, the output undergoes a feed-forward network (FFN) for non-linear transformation and feature extraction. Finally, the FFN output is added to the previously normalized result and undergoes an additional layer of normalization. The one-dimensional vector is then reshaped back into its 2D form, F5, for further processing in the subsequent network. The specific process can be described by Equations (1) and (2).

$$Q = K = V = Flatten(Input) \tag{1}$$

$$Output = Reshape(FNN(MultiHead(Q,K,V))) \tag{2}$$

Here, Flatten refers to the flattening operation, Q, K, and V are the results of applying the flattening operation on the 2D image, MultiHead denotes the multi-head attention mechanism, Reshape stands for the reshaping operation, and FFN represents the feed-forward network operation.

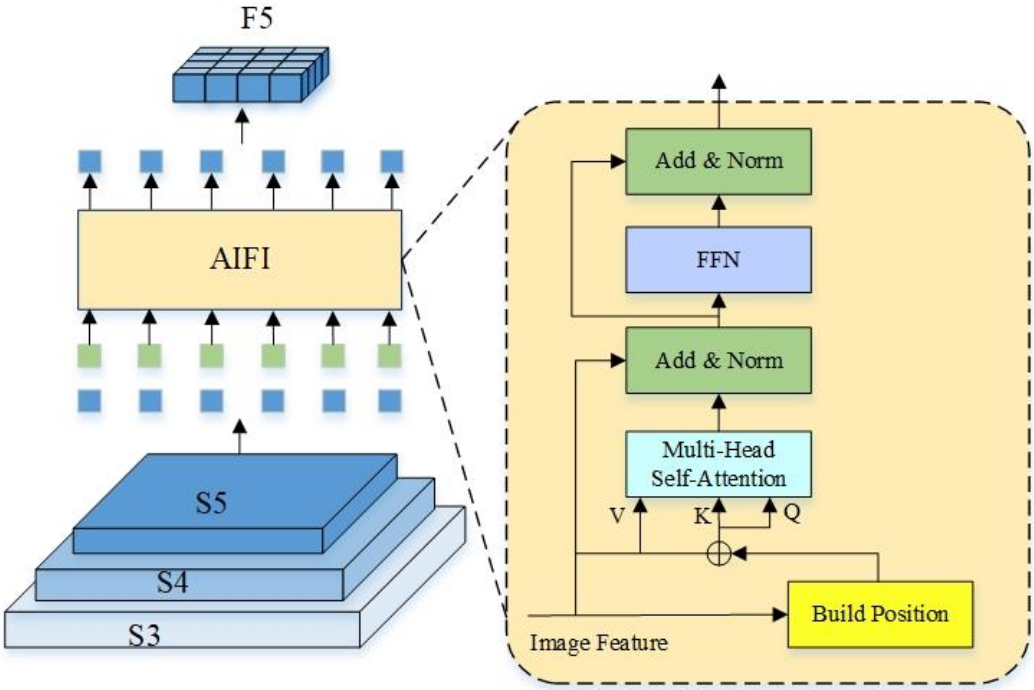

Fig. 5 - Structure of AIFI module, S3, S4, S5 represent different scale feature maps

Thanks to the Multi-Head Self-Attention and the FFN network, AIFI realize a scale-level interactions between advanced features, which helps the network better represent the connections of conceptual entities in the picture. This leads to improved extraction of subtle features of apple leaf spots, enhancing the detection performance and reducing false detections. Meanwhile, due to the feature fusion within the scale in AIFI, the computational cost of the detection model is reduced.

MPC Downsampling Enhancement Module

Downsampling techniques facilitate the processing of feature maps at different scales and objects by reducing the spatial size of feature maps. However, they also lead to information loss and a reduction in resolution. To address the partial loss of leaf spot information during downsampling, this study designs an improved downsampling module called MPC, as visualized in Fig. 6. The MPC downsampling is designed to strengthen the model's concentration on small lesion details in complex backgrounds, effectively preserving crucial contextual information and improving detection accuracy. The main components of the MPC are a 1×1 convolution, a Maxpool2d operation, a PConv, and a 3×3 convolution.

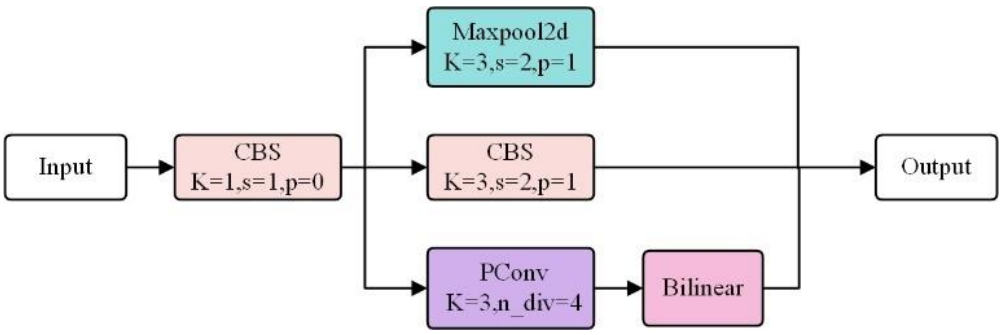

Fig. 6 - Structure of MPC module

Traditional downsampling in YOLOv8 uses a standalone 3×3 convolution module. While this module captures key features from the input data through filtering operations on the feature map, it also reduces the resolution of the feature map, impairing the capability to capture subtle patterns. To overcome this difficulty and improve the model's effectiveness and lightweight design, this study integrates PConv into the downsampling process. A comparison of the convolution operations between partial convolution and standard convolution is shown in Fig. 7.

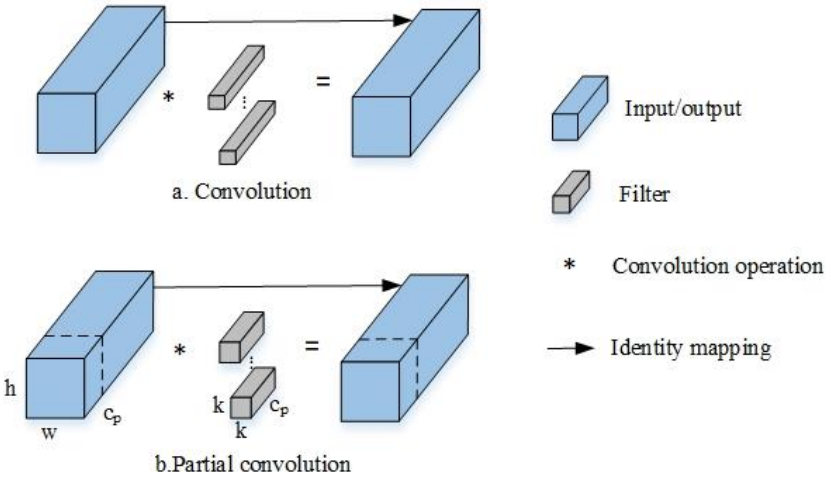

Fig. 7 - Convolutional operation comparison diagram

PConv leverages the inherent redundancy of feature representations by selectively performing standard convolution on a portion of the input channels without influencing the transformations of the remaining channels *(Fu et al., 2024)*. Here, h, w represent the input height and width dimensions, respectively; c is input channels count; c_p refers to used convolution channels count; k denotes the kernel size used for the partial convolution; and r indicates the ratio of used convolution channels. The Floating Point Operations (FLOPs) after using partial convolution can be represented by Equation (3).

$$F_{PConv} = h \times w \times k^2 \times {c_p}^2 \tag{3}$$

$$r = \frac{c_p}{c} \tag{4}$$

Given the default participation ratio r set to 1/4, PConv achieves only 1/16 of the computational complexity compared to standard convolution, significantly reducing the time and memory required for convolution operations.

During the downsampling enhancement process, MPC integrates PConv with Maxpool2d to better balance computational efficiency and information integrity. Partial convolution extracts spatial features by applying convolution operations to only a portion of channels, keeping the other channels unaltered during feature processing. This strategy enables the model to efficiently process complex input images while demonstrating excellent capability in extracting disease spot features in complex backgrounds. Furthermore, it keeps the model from obsessively concentrating on irrelevant information, such as the background, thus reducing unnecessary computational overhead. As a result, both detection precision and computing efficiency in leaf spot detection are improved.

Efficient Head

Compared to the coupled structure of the detection head in the YOLOv5 model, the YOLOv8 head design employs a decoupled structure, in which classification and regression operation are processed separately, as presented in Fig. 8. Specifically, each branch comprises a 3×3 convolutional and a 1×1 convolutional, with each branch designed to focus on its respective task.

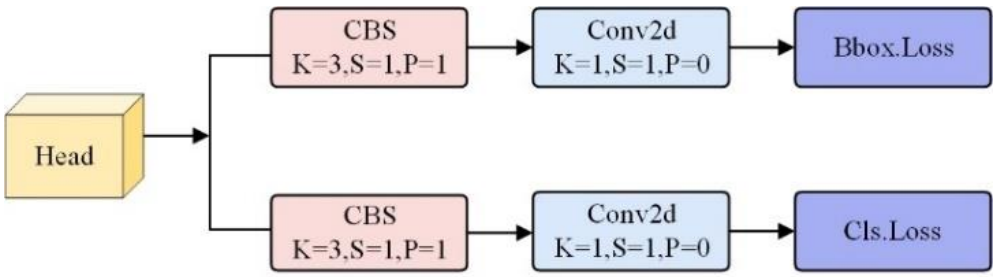

Fig. 8 - Structure of YOLOv8 Detection Head

To lessen the computational overhead and the parameter quantity while improving both detection speed and accuracy, this paper redesigns the Efficient Head based on the concept of parameter sharing, as shown in Fig. 9. The idea of merging first and then splitting is adopted, replacing the original two 3×3 convolutional blocks with a combination of the fast and efficient PConv and a 1×1 convolution. This improvement lowers parameter quantity and computing load while enabling more efficient feature extraction.

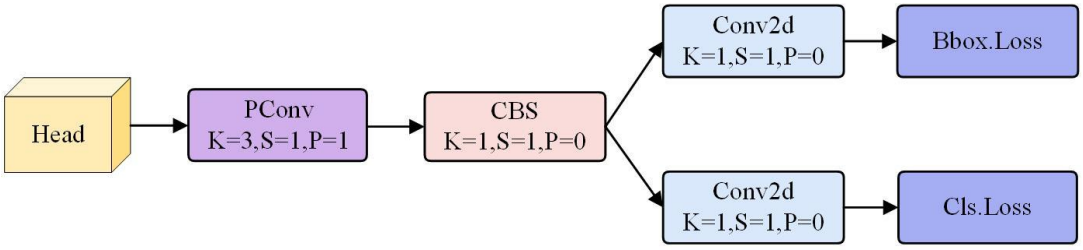

Fig. 9 - Structure of Efficient Head

RESULTS

Experimental Environment

The experimental environment comprised CUDA 11.7 on Ubuntu 18.04, using PyCharm as the development platform, PyTorch 2.0. as the deep learning framework, Python 3.8.10. An NVIDIA RTX 3060 GPU (12 GB) provided hardware acceleration. The training process spanned 150 epochs, adopting the Stochastic Gradient Descent (SGD) optimizer, an initial learning rate of 0.01, a batch size of 16, an initial learning rate of 0.01, with a patience parameter of 50.

Evaluation Metrics

To evaluate the model's performance impartially, frequent object detection metrics are employed, including Precision (P), Recall (R), mean Average Precision (mAP), and FLOPs. The formula is used for calculating P and R can be expressed by Equations (5) and (6).

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

Here, TP represents the quantity of diseases that the algorithm successfully recognized; FP represents the quantity of diseases that the algorithm incorrectly identified; FN represents the quantity of diseases not identified.

Although precision and recall are commonly used performance metrics, there is an inherent trade-off between them—optimizing one metric often comes at the expense of the other. Therefore, relying solely on either precision or recall fails to present a thorough and intuitive assessment of overall performance of a model. In contrast, mAP integrates characteristics of both metrics by calculating the average precision across varying recall levels, offering a more holistic evaluation. Furthermore, mAP more accurately reflects a model's performance under different detection difficulty levels, making it a more representative metric. The expression used to compute mAP is provided below:

$$AP = \int_0^1 P(R)dR \tag{7}$$

$$mAP = \frac{1}{c}\sum_{i=1}^{c} AP_i \quad (8)$$

Here c represents the count of classes in the dataset, i stands for the class index, and AP represents the area under the P-R curve for a single class.

Ablation Experiment

To evaluate the effectiveness of the suggested improvements in apple leaf lesion detection, ablation experiments were conducted on each improved method, as illustrated Table 1. All experiments were conducted without the use of transfer learning to ensure a fair and accurate evaluation of the proposed model's performance. A " √ " marks that the improved method was applied in the experimental group.

Table 1

Results of ablation experiments

No.	C2f-PE	AIFI	MPC	Efficient Head	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	FLOPs/G
1	×	×	×	×	90.5	80.8	89.9	55.0	28.4
2	√	×	×	×	90.1	81.1	90.9	55.7	33.3
3	×	√	×	×	91.5	81.3	90.9	56.1	28.3
4	×	×	√	×	90.1	81.2	90.3	55.3	29.4
5	×	×	×	√	89.8	81.7	90.5	55.2	21.5
6	√	√	×	×	89.6	82.4	91.1	56.2	33.2
7	√	√	√	×	91.6	82.0	91.2	56.1	34.1
8	√	√	√	√	91.3	82.3	91.4	56.4	27.2

In Experiment 2, the C2f-PE module was substituted for the C2f module, which leads to a 1.0% improvement in mAP@0.5 relative to the Experiment 1. This result confirms that the C2f-PE module can strengthen the model's capacity for localizing small disease spots, but it also increases the model's computational cost by 14.7%. Experiment 3 replaced the SPPF layer with the AIFI module, resulting in improvements of 0.5% in recall, 1.0% in mAP@0.5, and 1.1% in mAP@0.5:0.95. This demonstrates that the model benefits from intra-scale feature interaction, which enables better extraction of fine-grained features of apple leaf spots and enhances overall detection accuracy. In Experiment 4, the downsampling enhancement module MPC was added, resulting in a 0.4% improvement in mAP@0.5. This confirms that incorporating the MPC module can help the feature extraction process better preserve contextual information and make up for the information lost due to downsampling. After the Efficient Head was introduced in Experiment 5, the model's average precision improved by 0.6%, while its floating-point operations were reduced by 24.3%. This verifies the efficiency and lightweight characteristics of the Efficient Head. In Experiment 7, the ablation results demonstrated a 1.3% improvement in mAP@0.5 and 1.1% and 1.2% increases in precision and recall, respectively, compared with the baseline model. This improved detection performance but also increased the computational cost by 16.7%. Finally, the PAME-YOLO algorithm, which integrates four improvement methods, was compared to the baseline model. The improved PAME-YOLO demonstrated performance gains of 0.8%, 1.5%, 1.5%, and 1.4% in Precision, Recall, mAP@0.5, and mAP@0.5:0.95, respectively, while achieving a 4.2% reduction in computational load. Through a succession of experiments and comparative analysis, the improvements suggested in this paper has been effectively validated.

Detection Head Comparison

This study designs a novel lightweight detection head to enhance the detection efficiency and precision. In comparison experiments, several schemes were tested: (a) sharing two 3×3 convolutions; (b) sharing two 3×3 grouped convolutions; (c) sharing one 1×1 convolution and one 3×3 convolution; (d) sharing one PConv and one 1×1 convolution. Table 2 illustrates the comparison results of four schemes.

Table 2

Comparison experiment of different detection heads

Plan	P/%	R/%	mAP@0.5/%	FLOPs/G	Parameters/M
(a)	91.5	81.9	91.0	37.5	18246403
(b)	89.9	82.9	91.0	27.0	12311299
(c)	91.2	82.2	91.2	32.5	15493891
(d)	91.3	82.3	91.4	27.2	12589955

From the Table, it can be seen that although schemes (a) and (c) achieve higher mAP values, their computational cost and parameter count are too large, resulting in lower detection efficiency. Scheme (b), while exhibiting lower computational costs and fewer parameters, achieves the lowest precision among all schemes. Scheme (d) effectively balances detection precision with model complexity, delivering a high mAP value alongside a reduction in model parameters. Therefore, scheme (d) is selected as the detection head for the improved model to enhance detection efficiency.

Comparison with Current Advanced Algorithms

A comparison between the PAME-YOLO and other mainstream methods is conducted under the same condition and the same dataset. As presented in Table 3, the PAME-YOLO outperforms other mainstream object detection algorithms in terms of recall, mAP@0.5, and mAP@0.5:0.95. Specifically, the recall is higher than RT-DETR-L, YOLOv5s, YOLOv5m, YOLOv7-tiny, YOLOv8n, YOLOv8s, and YOLOv10s by 6.2%, 3.6%, 2.6%, 2.1%, 4.3%, 1.5%, and 0.4%, respectively, while the mAP@0.5 is higher by 5.2%, 3.0%, 2.0%, 3.4%, 2.8%, 1.5%, and 0.7%, respectively. The YOLOv5m model achieves the highest detection precision, but its parameter and computational requirements are too large. This means it requires higher computational resources and larger storage space to operate, making it unsuitable for real-time tasks. Considering all metrics, PAME-YOLO demonstrates higher detection precision, recall, and stability than other algorithms, with a reasonable model size and computational complexity, allowing it more appropriate for agricultural applications.

Table 3

Model comparison experiments

Model	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Parameters/M	FLOPs/G
RT-DETR-L	87.1	76.1	86.2	51.9	31989905	103.4
YOLOv5s	91.5	78.7	88.4	54.0	7018216	15.8
YOLOv5m	92.3	79.7	89.4	54.5	20861016	47.9
YOLOv7-tiny	87.2	80.2	88.0	51.6	6020400	13.2
YOLOv8n	89.9	78.0	88.6	54.2	3006233	8.1
YOLOv8s	90.5	80.8	89.9	55.0	11126745	28.4
YOLOv10s	90.6	81.9	90.7	54.5	8037282	24.5
PAME-YOLO	91.3	82.3	91.4	56.4	12589955	27.2

Comparison of different diseases under YOLOv8s and PAME-YOLO

A comprehensive comparative analysis of precision, recall, and mAP@0.5 was conducted for each type of leaf disease using the YOLOv8s and PAME-YOLO models. In Table 4, PAME-YOLO achieves notable improvements in all three metrics relative to the original YOLOv8s. This implies that the model possesses a stronger ability to recognize diseases and can effectively reduce the occurrence of missed detections and false positives, particularly for small target lesions.

Table 4

Comparison of different diseases under YOLOv8s and PAME-YOLO models

Class	YOLOv8s			PAME-YOLO		
	P/%	R/%	mAP@0.5/%	P/%	R/%	mAP@0.5/%
Alternaria leaf spot	91.6	69.7	85.8	92.4	71.2	86.7
Grey spot	89.1	82.4	88.4	89.5	85.1	91.6
Rust	90.8	90.4	95.5	92.0	90.5	95.9

Heatmap Visualization

In the apple leaf disease detection task, the Grad-CAM *(Selvaraju et al., 2017)* method was used to visually highlight the regions of interest and the areas where the model concentrates its attention during object detection, further enhancing the comprehension of the decision-making process. In the heatmap, darker pixels indicate a greater contribution to the prediction result, while lighter pixels indicate a smaller contribution. The heatmaps before and after model improvement are displayed in Fig. 10. It is evident that the PAME-YOLO model pays less attention to irrelevant information such as the background, and focuses more on the disease spots, making it better suited for complex orchard environments.

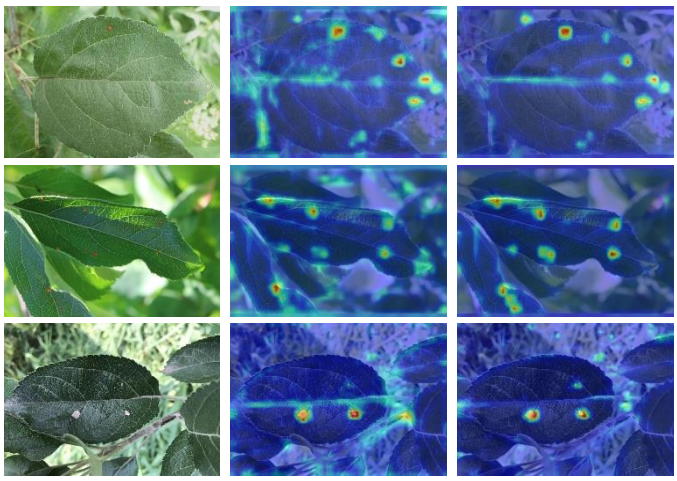

Fig. 10 - Visualization results of a heatmap

Visualization of Detection Results

Several images of leaves affected by Alternaria leaf spot, grey spot, and rust diseases were randomly selected to present detection results. As illustrated in Fig. 11, YOLOv8s shows suboptimal performance in detecting small spots on leaves in complex backgrounds. Specifically, it fails to detect some instances of Alternaria leaf spot and rust disease when the targets are small. In contrast, PAME-YOLO enhances the localization capability for small lesions by introducing the C2f-PE module, enabling accurate detection of spots without missed cases. Additionally, when detecting grey spot disease, the baseline model mistook the photographer's finger at the bottom left of the image for a lesion, resulting in a false detection. In contrast, PAME-YOLO correctly distinguished the actual spots from irrelevant objects, avoiding false detections and demonstrating higher robustness. In the figure, the yellow box represents a missed target, the blue box represents an error detection target.

Fig. 11 - Detection effect comparison diagram

Model Deployment and Application Implementation

As illustrated in Fig. 12, the trained model for apple leaf disease spot detection was quantized using the NCNN inference framework and it was deployed to a mobile platform, based on which an application named Apple Leaf Disease Detector was subsequently developed. To ensure cross-platform compatibility, the user interface was developed using uni-app. The application was then compiled into an APK file via Android Studio and subsequently installed and tested on a smartphone running Android 13. The main interface of the application includes an image preview panel, a button for uploading images from the gallery, a camera button for real-time capture, and a detection button.

Users are provided with the option to either import apple leaf images from the device gallery or acquire them in real time via the built-in camera. After the image is uploaded or captured, the user can click the "Detection" button to initiate model inference. The detection results are displayed on the image in the form of bounding boxes, along with disease category labels and confidence scores. As can be observed from Fig. 13, this app can accurately detect apple leaf lesion across different categories even in challenging orchard environments.

The deployed application is lightweight and responsive, supporting accelerated inference via NCCN. This enables efficient, low-latency identification of apple leaf diseases in the field, offering strong practical value for fruit growers.

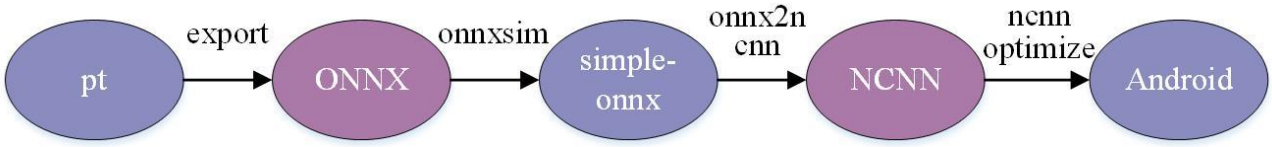

Fig. 12 - Model quantization and conversion pipeline

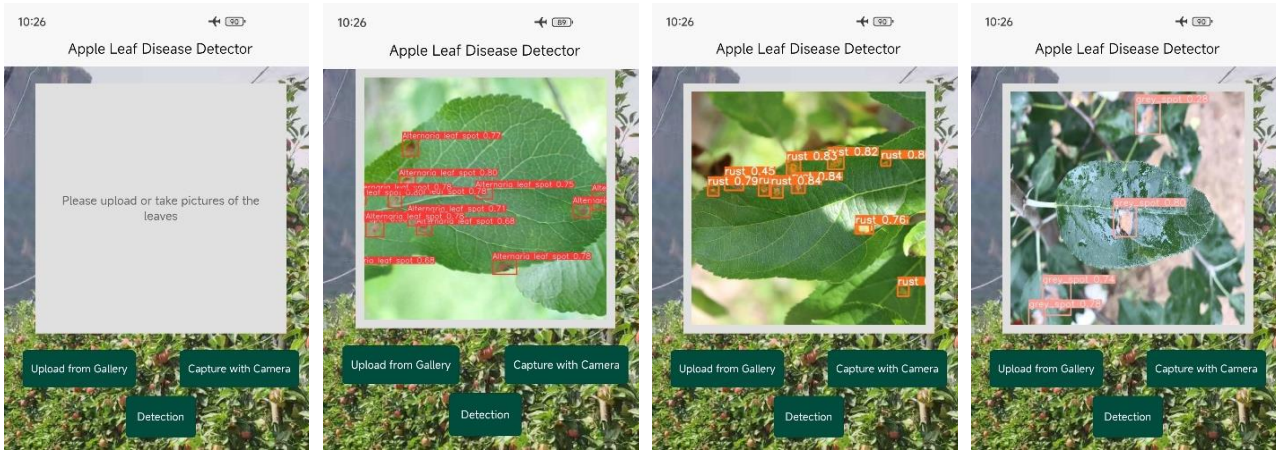

Fig. 13 - Detection of Apple Leaf Diseases on Mobile Platforms

CONCLUSIONS

This study proposed an enhanced object detection algorithm called PAME-YOLO that focused on the issues of low detection accuracy and the susceptibility to overlooking small lesion targets in apple leaf disease diagnosis under complicated backgrounds. Specifically, the C2f-PE feature extraction module was designed to improve the model's detection accuracy for small lesion targets. In addition, an intra-scale feature interaction mechanism was introduced to capture more fine-grained lesion information and reduce false detections. The downsampling enhancement module, MPC, was designed to ensure that critical contextual information was comprehensively preserved at the feature extraction stage. Lastly, a lightweight and efficient detection head was employed to reduce model parameters and computational cost, thereby enhancing both detection efficiency and accuracy. The outcomes of the experiment demonstrate that, compared with the YOLOv8s baseline, the improved algorithm yields an increase of 1.5% in recall, 1.5% in mAP@0.5, and 1.4% in mAP@0.5:0.95. Relative to other mainstream algorithms, the proposed algorithm shows exceptional detection performance under complicated conditions, highlighting its advantages for practical applications. This provides important technical assistance for the early control and management of apple leaf diseases, helping to reduce crop loss and improve orchard health. In subsequent research, the dataset will be further expanded by incorporating a greater diversity of apple leaf disease images, aiming to strengthen the model's generalization capability and practical usefulness.

ACKNOWLEDGEMENT

The author has been supported by the National Social Science Foundation of China (No. 24BTQ031).

REFERENCES

[1] Abulizi, A., Ye, J., Abudukelimu, H., Guo, W. (2024). DM-YOLO: improved YOLOv9model for tomato leaf disease detection. *Frontiers in Plant Science*, 15, 1473928. DOI: https://doi.org/10.3389/fpls.2024.1473928.

[2] Bai, X., Li, Z., Li, W., Zhao, Y., Li, M., Chen, H., et al. (2021). Comparison of Machine-Learning and CASA Models for Predicting Apple Fruit Yields from Time-Series Planet Imageries. *Remote Sensing*, 13(16), 3073. DOI: https://doi.org/10.3390/rs13163073.

[3] Bi, J., Li, K., Zheng, X., Zhang, G., Lei, T. (2025). SPDC-YOLO: An Efficient Small Target Detection Network Based on Improved YOLOv8 for Drone Aerial Image. *Remote Sensing*, 17(4), 685. DOI: https://doi.org/10.3390/rs17040685.

[4] Chen, J., Kao, S., He, H., Zhuo, W., Wen, S., Lee, C. (2023). Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. *In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.12021-12031, Vancouver, BC, Canada. DOI: https://doi.org/10.1109/CVPR52729.2023.01157.

[5] Fu, C., Ren, L., Wang, F. (2024). Recognizing beef cattle behavior under automatic scene distinction using lightweight FABF-YOLOv8s (自动化场景区分下 FABF-YOLOv8s 轻量化肉牛行为识别方法). *Transactions of the Chinese Society of Agricultural Engineering*, 40(15), 152-163. DOI: https://doi.org/10.11975/j.issn.1002-6819.202404073.

[6] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *In 2014 IEEE Conference on Computer Vision and Pattern Recognition,* pp.580-587, Columbus, OH, USA. DOI: https://doi.org/10.1109/CVPR.2014.81.

[7] Gong, X., Zhang, S. (2023). A High-Precision Detection Method of Apple Leaf Diseases Using Improved Faster R-CNN. *Agriculture*, 13(2), 240. DOI: https://doi.org/10.3390/agriculture13020240.

[8] Gao, L., Zhao, X., Yue, X., Yue, Y., Wang, X., Wu, H., et al. (2024). A Lightweight YOLOv8 Model for Apple Leaf Disease Detection. *Applied Sciences*, 14(15), 6710. DOI: https://doi.org/10.3390/app14156710.

[9] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. DOI: https://doi.org/10.1038/nature14539.

[10] Li, H., Shi, L., Fang, S., Yin, F. (2023). Real-Time Detection of Apple Leaf Diseases in Natural Scenes Based on YOLOv5. *Agriculture*, 13(4), 878. DOI: https://doi.org/10.3390/agriculture13040878.

[11] Ouyang, D., H, S., Zhang, G., Luo, M., Guo, H., Zhan, J. (2023). Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Pro-cessing (ICASSP)*, pp.1-5, Rhodes Island, Greece. DOI: https://doi.org/10.1109/ICASSP49357.2023.10096516.

[12] Selvaraju, RR., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *In 2017 IEEE International Conference on Computer Vision (ICCV),* pp.618-626, Venice, Italy. DOI: https://doi.org/10.1109/ICCV.2017.74.

[13] Tian, Y., Zhao, C., Zhang, T., Wu, H., Zhao, Y. (2024). Recognition Method of Cabbage Heads at Harvest Stage under Complex Background Based on Improved YOLOv8n. *Agriculture*, 14(7), 1125. DOI: https://doi.org/10.3390/agriculture14071125.

[14] Wang, H., Zhang, Y., Zhu, C. (2025). DAFPN-YOLO: An Improved UAV-Based Object Detection Algorithm Based on YOLOv8s. *Computers, Materials & Continua*, 83(2), 1929-1949. DOI: https://doi.org/10.32604/cmc.2025.061363.

[15] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.11531-11539, Seattle, WA, USA. DOI: https://doi.org/10.1109/CVPR42600.2020.01155.

[16] Woo, S., Park, J., Lee, JY., Kweon, IS. (2018). CBAM: Convolutional Block Attention Module. *In Proceedings of the European conference on computer vision (ECCV)*, p.3-9, Munich, Germany. DOI: https://doi.org/10.1007/978-3-030-01234-2_1.

[17] Xu, S., Zheng, S., Xu, W., Xu, R., Wang, C., Zhang, J., et al. (2024). HCF-Net: Hierarchical context fusion network for infrared small object detection. *In 2024 IEEE International Conference on Multimedia and Expo (ICME)*, pp.1-6, Niagara Falls, ON, Canada. DOI: https://doi.org/10.1109/ICME57554.2024.10687431.

[18] Yang, Q., Duan, S., Wang, L. (2022). Efficient Identification of Apple Leaf Diseases in the Wild Using Convolutional Neural Networks. *Agronomy*, 12 (11), 2784. DOI: https://doi.org/10.3390/agronomy12112784.

[19] Zhang, K., Wu, Q., Chen, Y. (2021). Detecting soybean leaf disease from synthetic image using multi-feature fusion faster R-CNN. *Computers and Electronics in Agriculture*, 183, 106064. DOI: https://doi.org/10.1016/j.compag.2021.106064.

[20] Zhao, y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q. (2024). DETRs Beat YOLOs on Real-time Object Detection. *In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.16965-16974, Seattle, WA, USA. DOI: https://doi.org/10.1109/CVPR52733.2024.01605.