RESEARCH ON GRAPE DISEASE DETECTION METHOD BASED ON PWE-YOLO
/
基于 PWE-YOLO 的葡萄病害检测方法

Haoyue LIU [1], Benzhi YANG [1], Mingyang Gao [1], Dong WANG [1], Yuepeng SONG[1,2,*] Longlong REN[1,2,*]
[1] Shandong Agricultural University, College of Mechanical and Electrical Engineering/ China;
[2] Shandong Provincial Engineering Laboratory of Agricultural Equipment Intelligence/ China
E-mail: uptonsong@sdau.edu.cn
DOI: https://doi.org/10.35633/inmateh-79-40

Keywords: *Grape diseases; Deep learning; Object detection; Lightweight model; YOLO v11*

ABSTRACT
Accurate detection of grape diseases in controlled facility environments is of significant importance. In this study, a novel detection model, PWE-YOLO, is proposed based on YOLO v11. In this model, the C3k2 modules in the Neck network are replaced with C3k2-PConv modules, which maintain high-precision feature extraction while reducing model parameters and memory consumption. Furthermore, a WaveletPool module is introduced to replace all Conv modules in the model, except for the 0th and 1st layers, enhancing feature representation and detection accuracy for grape diseases while further reducing model parameters and floating-point operations. EMA attention modules are incorporated at the 17th and 24th layers to improve small-object detection capabilities, thereby increasing overall detection precision. Experimental results demonstrate that PWE-YOLO achieves a precision of 85.0%, recall of 83.6%, and mAP of 86.9%, with 1.96 × 10^6 parameters and 5.1 × 10^9 floating-point operations. Compared with the original YOLO v11 model, precision, recall, and mAP increase by 4.5, 1.4, and 1.9 percentage points, respectively, while parameters and floating-point operations decrease by 24.03% and 10.05%. Relative to YOLO v8, v9, v10, v12, and v13, precision improves by 2.2%, 5.3%, 3.3%, 2.2%, and 2.5%, recall by 6.0%, 2.8%, 5.2%, 8.8% and 3.8%, and mAP by 1.9%, 1.8%, 2.8%, 5.0%, and 2.3%, respectively, with corresponding reductions in model parameters and floating-point operations. These results indicate that PWE-YOLO not only provides high-accuracy detection for grape diseases but also reduces computational complexity, achieving a lightweight architecture and faster detection speed, making it suitable for deployment in resource-constrained target detection scenarios.

摘要
为实现设施环境中葡萄病害目标的准确检测，本文在 YOLO v11 的基础上，提出一种适用于设施环境葡萄病害目标检测模型 PWE-YOLO。将颈部网络中的 C3k2 模块替换为 C3k2-PConv 模块，在保持高精度特征提取能力的基础上，降低模型参数量和内存占用量；同时在模型中引入 WaveletPool 模块，替换模型中除第 0 层、第 1 层外的所有 Conv 模块，提高葡萄病害的特征表达能力与检测精度，进一步降低模型参数量和浮点运算量；且在模型的第 17 层和第 24 层中添加 EMA 注意力机制模块，增强模型对小目标的检测能力，提升模型的检测精度。结果显示，PWE-YOLO 模型的精确率为 85.0%、召回率为 83.6%、mAP 为 86.9%、参数量为 1.96×10^6、浮点运算量为 5.1×10^9。与原始 YOLO v11 模型相比，该模型的精确率、召回率和 mAP 分别提高 4.5、1.4、1.9 个百分点，参数量和浮点运算量分别降低 24.03%、10.05%。与 YOLO v8、YOLO v9、YOLO v10、YOLO v12、和 YOLO v13 模型相比，其精确率分别提高 2.2、5.3、3.3、2.2、和 2.5 个百分点，召回率分别提高 6.0、2.8、5.2、8.8、和 3.8 个百分点，mAP 分别提高 1.9、1.8、2.8、5.0、和 2.3 个百分比，参数量分别减少 34.88%、25.19%、27.40%、21.91%和 20.00%，浮点运算量分别减少 37.03%、52.34%、37.80%、12.07%和 17.74%。结果表明，本文提出的 PWE-YOLO 模型在葡萄病害检测任务中不仅能够提供高准确性的检测性能，同时也有效降低了计算复杂度，具有较轻的模型结构和较快的检测速度，适合应用于对计算资源要求较高的目标检测任务中。

INTRODUCTION
Grapes, as one of the most important economic crops worldwide, are known to be highly susceptible to fungal diseases, including downy mildew, powdery mildew, and gray mold, during their growth period. Premature leaf senescence and fruit rot are caused by these diseases, which severely affect yield and quality *(Rahman et al, 2024)*.

Haoyue Liu, M.Sc. Stud. Eng.; Benzhi Yang, M.Sc. Stud. Eng.; Mingyang Gao, M.Sc. Stud. Eng.; Dong Wang, M.Sc. Stud. Eng.; Longlong Ren Lecturer Ph.D. Eng; Yuepeng Song, Prof. Ph.D. Eng.

Therefore, early detection and effective control of these diseases are considered crucial for ensuring the sustainable development of the grape industry. However, in the early stages of infection, disease symptoms are often subtle and typically manifest as small spots or slight discoloration on leaves, which can be easily mistaken for leaf changes induced by environmental stress. Traditional manual identification methods are largely dependent on expert judgment, resulting in relatively low accuracy and making comprehensive monitoring in large-scale vineyards difficult to achieve *(Poblete-Echeverría et al, 2025)*.

In recent years, with the rapid advancement of computer vision and deep learning technologies, image recognition methods based on convolutional neural networks (CNNs) have made remarkable progress in crop disease detection *(Zhang et al, 2022)*. Among these methods, the YOLO series of algorithms has been extensively applied in agricultural settings, particularly for the identification and detection of grape diseases, due to its end-to-end architecture and high real-time performance. For example, Liu et al. proposed an improved CNN-based detection framework for multiple grape leaf diseases, in which preprocessing and data augmentation techniques were applied to simulate real-world environments, enhancing the model's robustness and generalization ability, and achieving a precision of 97.22% on the test set *(Liu et al, 2020)*. Zhen Yuanyuan et al. introduced the CBAM attention mechanism on top of the YOLOv8 model to enhance the feature representation of key disease spots on grape leaves, while the ShuffleNet V2 module was incorporated for network lightweight modification, which improved both detection accuracy and speed, achieving a mAP of 94.6% and an inference speed of 52.06 FPS *(Yang et al, 2024)*. Zhang et al. proposed the DLVTNet model, combining generative data augmentation with a lightweight Transformer, and achieved a precision of 98.48% on the grape leaf disease dataset *(ZHEN et al, 2025)*. Fu et al. combined YOLO and Transformer concepts to develop the MHDI-DETR model, balancing computational efficiency and detection accuracy, and achieving a mAP of 92.6% in detecting grape leaf diseases such as black rot and black spot *(Fu et al, 2024)*. Zhang Lin et al. proposed a detection method combining StyleGAN2-ADA augmentation with an improved YOLOv7 model to enhance localization of small early-stage disease spots, and by utilizing attention mechanisms and feature fusion, the model was able to detect small targets effectively, achieving a mAP of 95.4%, a precision of 94.1%, and an inference speed of 77.12 FPS *(Zhang et al, 2025)*.

Nevertheless, existing studies remain limited, as deep network models typically involve large numbers of parameters and high computational complexity, which hinder efficient deployment on resource-constrained agricultural edge devices. To address these challenges, an early detection method for grape diseases based on PWE-YOLO is proposed in this study. The proposed model enhances detection accuracy for grape diseases, reduces model parameters and memory consumption, and improves detection speed. Notably, in resource-limited environments, the model demonstrates superior real-time detection and on-site deployment capabilities, thereby exhibiting increased practicality and robustness.

MATERIALS AND METHODS

Data Collection and Preprocessing

Image data were collected at the Yantai Academy of Agricultural Sciences, Fushan District, Yantai City (37.5076°N, 121.255°E), from September 2024 to August 2025. The images were captured under natural lighting conditions using an iPhone 15 Pro smartphone, which took multi-angle shots of grape leaves. A total of approximately 3,000 images were collected, covering both healthy leaves and five major disease types: downy mildew, powdery mildew, gray mold, canker, and Grapevine Leafroll-associated Virus (GLRaV) disease. To ensure sample diversity, the images were captured under varying lighting intensities, humidity levels, and leaf age stages. The collected images were manually screened to remove blurred and occluded samples, after which they were resized to 600×400 pixels and normalized to ensure consistency in input for model training. After screening, a final dataset consisting of 1,785 high-quality images was created. Table 1 presents the number of images for each disease category in the dataset.

Table 1

Number of images in grape disease dataset

Type	downy mildew	powdery mildew	gray mold	canker	GLRaV
Primitive	456	359	321	369	280

Disease annotation was performed using the LabelImg tool, with the annotation categories including five types: downy mildew, powdery mildew, gray mold, canker, and GLRaV; the annotation results are shown in Figure 1.

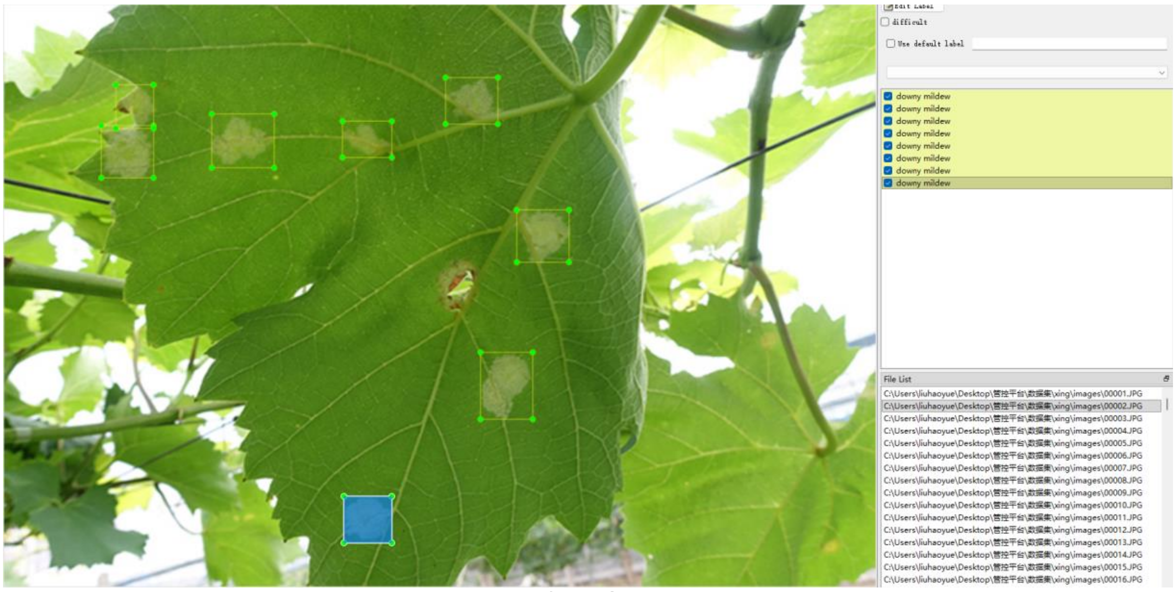

Fig. 1 - Annotation of the Grape Disease Dataset

The dataset was divided into training, validation, and test sets with a ratio of 70%, 20%, and 10%, respectively. To improve the model's robustness and generalization ability, data augmentation was applied to the images in the dataset. Specifically, six augmentation techniques were used: adding Gaussian noise, random rotation, random cropping, horizontal flipping, brightness adjustment, and Gaussian blur. For each original image, three techniques were randomly selected) from the six listed and applied in combination, repeated three times. This process produced three augmented samples per original image, effectively enlarging the dataset and improving the model's adaptability to complex imaging conditions. The dataset was ultimately expanded to 7,140 images, with 4,998 images in the training set, 1,428 in the validation set, and 714 in the test set. Representative examples of the augmented dataset images are shown in Figure 2.

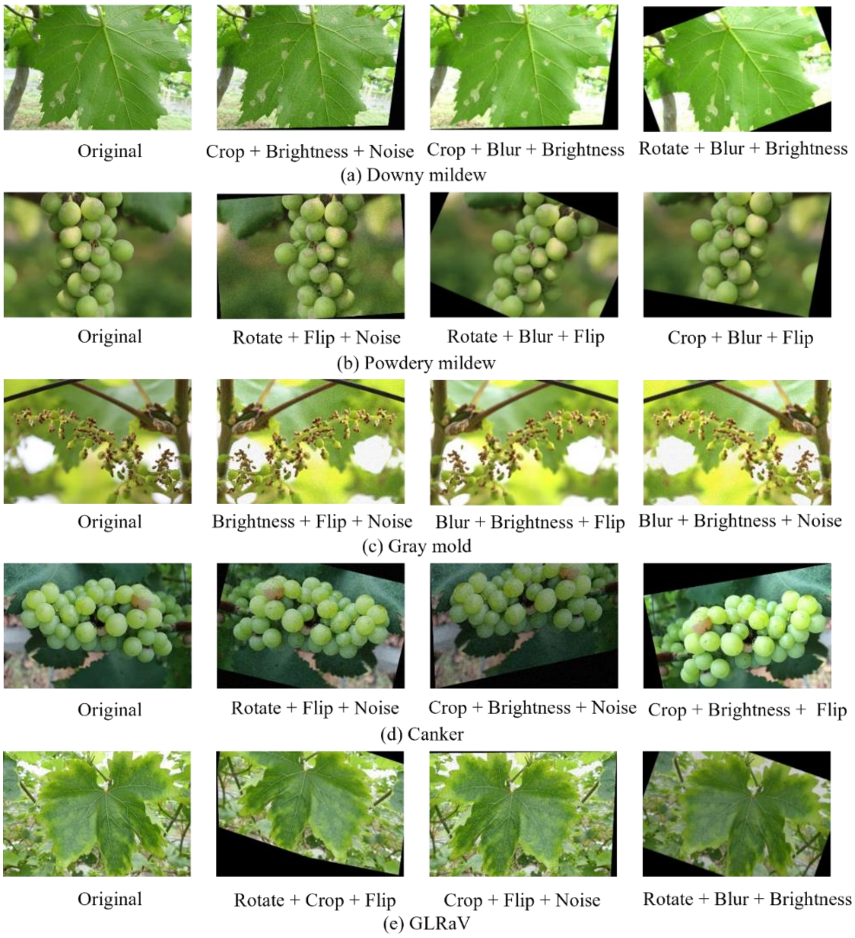

Fig. 2 - Image Examples of the Grape Disease Dataset

Construction of the PWE-YOLO-Based Grape Disease Detection Model

In agricultural object detection tasks, YOLO v11 adopts an efficient single-stage architecture and demonstrates strong small-object feature representation capabilities, enabling high-precision real-time detection of targets such as pests and diseases in complex scenarios, while maintaining a lightweight design and ease of deployment *(Badgujar et al, 2024)*. To enhance the detection performance of grape diseases in controlled environments, the YOLO v11 model was improved in this study.

First, the C3k2-PConv module was introduced to replace the C3k2 module in the Neck network. This modification reduces the model's parameter count and memory usage while preserving high-precision feature extraction capabilities. Consequently, real-time detection performance is improved, and model robustness and detection accuracy are simultaneously enhanced.

Second, the WaveletPool module was introduced to replace all Conv modules in the model, except for the 0th and 1st layers. This modification reduces the model's parameter count and floating-point operations, maintaining a lightweight architecture while effectively suppressing downsampling aliasing. Moreover, it enhances the retention of high-frequency features in disease regions, such as edges, patterns, and color variations, thereby improving feature representation and detection accuracy for early-stage grape diseases.

Finally, the EMA attention mechanism module was added to the 17th and 24th layers to mitigate background interference, enhance the model's robustness in complex environments, and improve precision and mAP values. The architecture of the proposed PWE-YOLO model is illustrated in Figure 3.

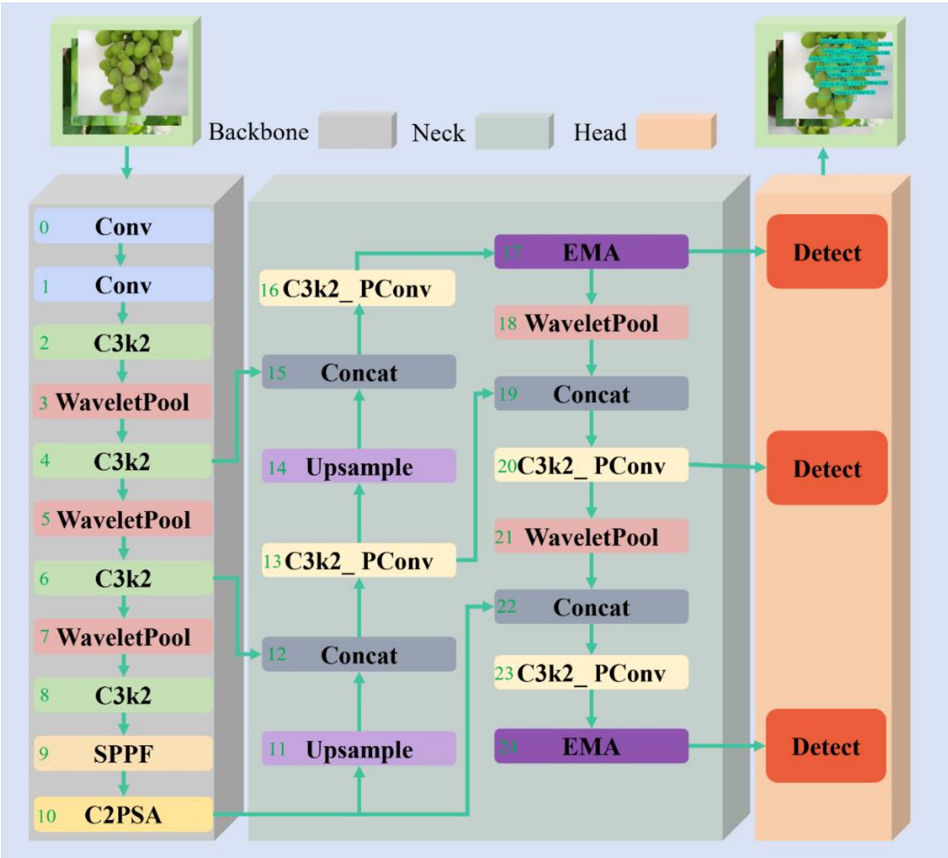

Fig. 3 - Architecture of the PWE-YOLO Model

C3k2-PConv Module

Partial Convolution (PConv) *(Chen et al, 2023)* performs convolution operations on the valid regions of the input feature map, reducing redundant computations and memory accesses, thereby improving the efficiency of spatial feature extraction *(Jiang et al, 2025)*. Unlike traditional convolution and depth/group convolutions, PConv applies the convolution kernel only to valid channels, leaving other channels unchanged. This approach reduces the number of floating-point operations, maintains high detection accuracy, reduces the model's parameter count, and increases computational speed *(Zhang and Zeng, 2024)*, The structure of PConv is illustrated in Figure 4. The C3k2 module extracts multidimensional features from the input image by combining convolutions at different levels. The C3 layer is generally used to capture feature information at multiple scales more effectively, thereby enhancing the model's ability to perceive complex image content *(Jin et al, 2025; Zhang et al, 2026)*.

The C3k2-PConv module combines the deep feature extraction capability of Convolutional Neural Networks (CNN) with the selective convolution ability of partial convolutions. Through flexible convolution kernel combinations and localized convolution operations, it enhances the processing of complex image content while effectively addressing the impact of missing or invalid data. The structure of the C3k2-PConv module is shown in Figure 5.

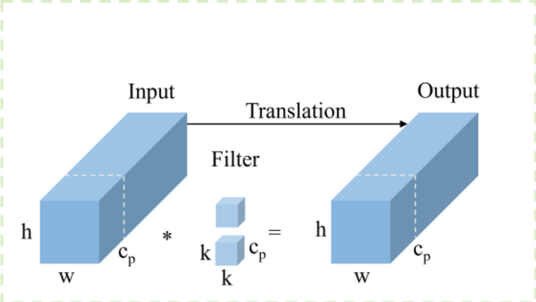

Fig. 4 - Architecture of the PConv Module

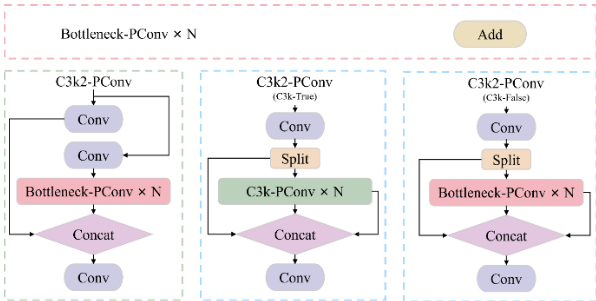

Fig. 5 - Architecture of the C3k2-PConv Module

For the early detection of grape diseases, the C3k2-PConv module substantially reduces the model's parameter count and memory usage while maintaining high-precision feature extraction capabilities, thereby enhancing real-time detection speed. Particularly in cases of feature loss or damaged regions caused by insufficient lighting or suboptimal shooting angles, C3k2-PConv performs convolution operations exclusively on valid regions, avoiding interference from missing or corrupted areas in the final detection results. Consequently, the module improves the model's robustness and detection accuracy under challenging imaging conditions.

WaveletPool Module

The WaveletPool module is an efficient downsampling method based on discrete wavelet transform *(Sun et al, 2022)*. By performing two-dimensional wavelet decomposition on the input feature map, the module extracts a low-frequency structural subband (LL) and three types of high-frequency detail subbands (LH, HL, HH), using the LL subband as the downsampling output. This approach reduces feature dimensionality while retaining key information such as edges and textures of the target *(Qu and Zhang, 2025; Xu and Wang, 2023)*. Compared with traditional max pooling or average pooling, WaveletPool module effectively suppresses aliasing in the downsampling process in the transform domain, preventing small target features from being overly smoothed or lost in deep networks. It is particularly suitable for small object recognition and detection tasks in agricultural scenarios characterized by significant scale variations and complex backgrounds *(Raj et al, 2025)*. Furthermore, WaveletPool features low computational cost and a lightweight structure, making it appropriate for deployment on mobile devices or in real-time detection systems. Its advantages in enhancing accuracy and feature stability have been validated in several lightweight detection networks. The structure of the WaveletPool module is shown in Figure 6.

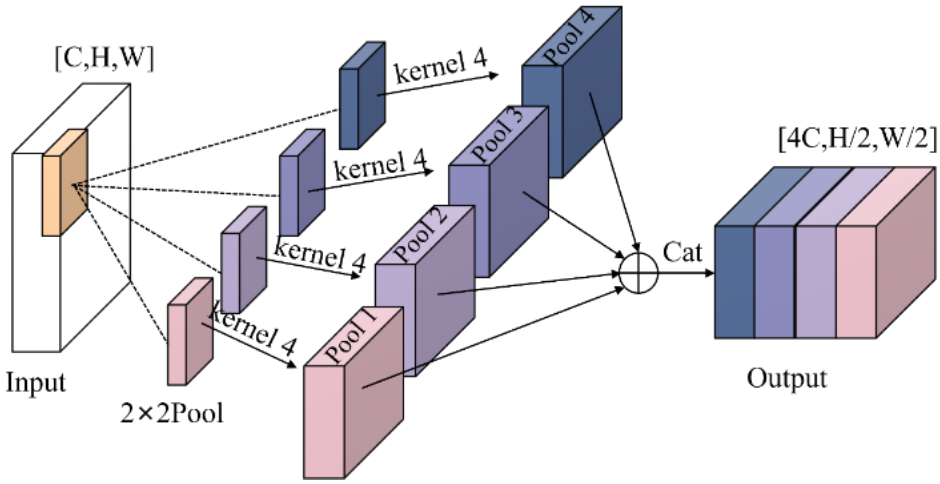

Fig. 6 - Architecture of the WaveletPool module

For grape disease detection tasks, the WaveletPool module, based on discrete wavelet transform, effectively suppresses aliasing during downsampling while maintaining the network's lightweight design, enhancing the retention of high-frequency features such as edges, patterns, and color variations in disease regions. This module stabilizes the model under conditions of low texture contrast, complex background lighting, and micro-scale variations in early-stage disease spots, thereby improving feature representation and detection accuracy for early-stage grape diseases.

EMA Attention Module

The EMA (Efficient Multi-scale Attention) *(Ouyang et al, 2023)* mechanism is an efficient feature extraction module designed to enhance the fusion of multi-scale features. Integrating the EMA attention mechanism into the YOLO model can improve the detection accuracy of agricultural targets, including crop diseases and small objects such as buds. The EMA module divides the input feature map into multiple sub-feature groups and utilizes parallel 1×1 and 3×3 convolution paths to extract fine-grained and large-scale features, respectively. The 1×1 path employs global average pooling and convolution operations to encode channel information, enhancing inter-channel interaction, while the 3×3 path captures broader spatial features, expanding the receptive field for feature extraction *(Li et al, 2025)*. This design allows the EMA attention mechanism to focus on target regions without compromising computational efficiency, improving robustness to small targets and complex backgrounds *(Williams et al, 2018)*. Following its integration, the YOLO model demonstrates enhanced capability in detecting small targets and handling complex backgrounds, thereby improving its suitability for agricultural application, effectively improving the detection of small objects such as buds and early-stage disease spots. The structure of the EMA attention mechanism is shown in Figure 7.

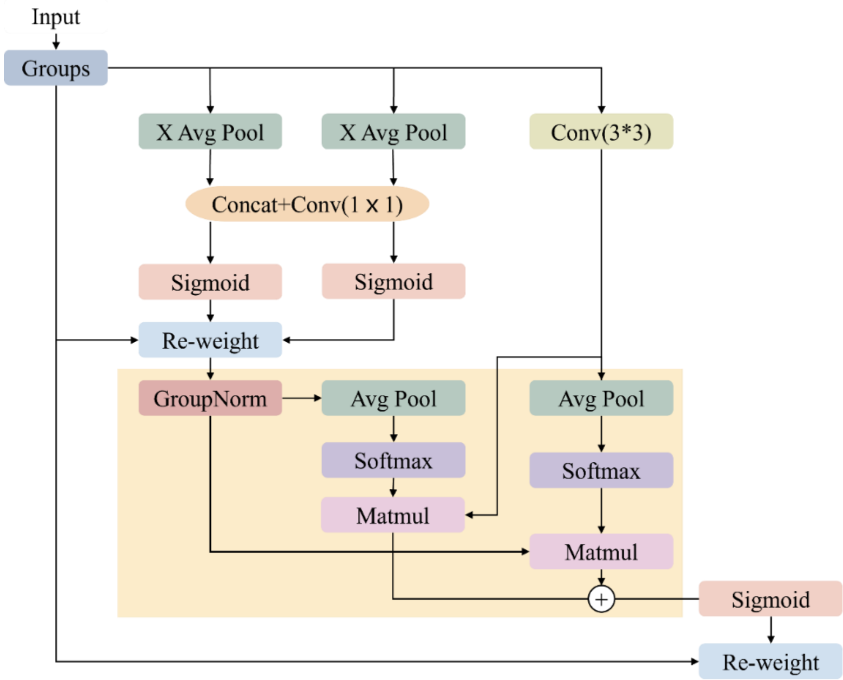

Fig. 7 - Architecture of the EMA module

In grape disease detection tasks, the EMA attention mechanism enhances the model's sensitivity to early-stage grape diseases, such as small lesions, through multi-scale feature extraction and information fusion. It effectively captures features at different scales via parallel 1×1 and 3×3 convolution paths, while reinforcing attention to small targets, thereby improving detection accuracy for small lesions. Additionally, EMA reduces background interference, enhancing model robustness in complex environments and enabling more accurate and stable early detection of grape diseases.

Model Deployment Methodology

The operating system used is Linux Ubuntu 22.04, with an Intel(R) Xeon(R) Platinum 8474C CPU model, an NVIDIA GeForce RTX 4090 GPU with 24GB of video memory. The programming language is Python 3.10, the deep learning framework is PyTorch 2.1.0, and the GPU acceleration library is CUDA 12.1.

RESULTS

Experimental Setup and Evaluation Metrics

To evaluate the performance of the network model in object detection tasks, Precision, Recall, mean Average Precision (mAP), the number of parameters, and the number of floating-point operations are used as the primary evaluation metrics *(Xie et al, 2025)*. During training, the input image size was set to 400 × 600 pixels, the number of epochs was set to 100, and the batch size was set to 64. The SGD optimizer was used for model optimization. The cache mechanism was enabled, the number of data loading workers was set to 8, and automatic mixed precision training was adopted.

Comparison Experiment of C3k2-PConv Module

To evaluate the performance of the C3k2-PConv module proposed in this study, alternative modules—C3k2-WTConv, C3k2-FasterBlock, and C3k2-RepVGG—were used to replace the corresponding module at the same position in the model, and comparison experiments were conducted. The results are summarized in Table 2.

Table 2

Effects of different C3k2 modules on detection performance

Model	P/%	R/%	mAP/%	Parameters	FLOPs
YOLO v11	80.5	82.2	85.0	$2.58{\times}10^6$	$6.3{\times}10^9$
YOLO v11-C3k2-PConv	81.5	82.7	85.7	$2.37{\times}10^6$	$5.9{\times}10^9$
YOLO v11-C3k2-WTConv	80.5	81.8	85.5	$2.53{\times}10^6$	$6.3{\times}10^9$
YOLO v11- C3k2-FasterBlock	82.0	82.5	85.1	$2.54{\times}10^6$	$6.2{\times}10^9$
YOLO v11-C3k2-RepVGG	80.9	82.5	85.2	$2.60{\times}10^6$	$6.4{\times}10^9$

As shown in the table, the YOLO v11-C3k2-PConv model achieves improvements of 1.0, 0.5, and 0.7 percentage points in Precision, Recall, and mAP, respectively, compared with the baseline YOLO v11 model, while reducing the number of parameters and floating-point operations by 8.1% and 6.3%, respectively. Compared with the other alternative modules, the YOLO v11-C3k2-PConv model exhibits slightly lower Precision than the YOLO v11-C3k2-FasterBlock model, but outperforms the remaining models across all other evaluation metrics. In particular, its Recall, mAP, number of parameters, and floating-point operations are superior to those of the alternative models, demonstrating a lower computational cost. These results suggest that the YOLO v11-C3k2-PConv model is well adapted to object detection tasks under constrained computational resources and demonstrates superior performance in the grape disease detection task evaluated in this study.

Comparison Experiment of WaveletPool Module

The performance of the WaveletPool module proposed in this study was evaluated by replacing it with alternative modules—SAConv, ContextGuided, and ADown—at the same position in the model, and comparison experiments were conducted. The results are summarized in Table 3.

Table 3

Comparison of detection results for different modules

Model	P/%	R/%	mAP/%	Parameters	FLOPs
YOLO v11	80.5	82.2	85.0	$2.58{\times}10^6$	$6.3{\times}10^9$
YOLO v11-WaveletPool	83.2	80.0	85.5	$2.17{\times}10^6$	$5.4{\times}10^9$
YOLO v11-SAConv	83.2	79.9	84.3	$3.42{\times}10^6$	$4.7{\times}10^9$
YOLO v11-ContextGuided	81.2	78.8	86.0	$2.83{\times}10^6$	$6.8{\times}10^9$
YOLO v11-ADown	83.0	80.0	83.7	$2.11{\times}10^6$	$5.3{\times}10^9$

As shown in the table, compared with the original YOLO v11 model, the YOLO v11-WaveletPool model exhibits a slight decrease in Recall, while Precision and mAP increase by 2.7 and 0.5 percentage points, respectively. Additionally, the number of parameters and floating-point operations decrease by 15.89% and 14.29%, respectively. Compared with the alternative modules, YOLO v11-WaveletPool demonstrates higher Precision and Recall. Although its mAP is slightly lower than that of the YOLO v11-ContextGuided model, it outperforms the other models in overall detection performance. In comparison with the YOLO v11-ADown model, the YOLO v11-WaveletPool model has slightly higher numbers of parameters and floating-point operations, but its Precision, Recall, and mAP remain superior. Moreover, the total number of parameters and floating-point operations in the YOLO v11-WaveletPool model is substantially lower than in the remaining models, contributing to better overall efficiency and detection performance.

Comparison Experiment of EMA Attention Module

The performance of the EMA attention mechanism proposed in this study was evaluated by integrating alternative attention mechanisms—including SCSA, MSDA, LSKA, iRMB, CGA, and SEAM—at the corresponding positions in the model. Performance comparisons were then conducted, and the results are presented in Figure 8.

As shown in Figure 8, compared with the baseline YOLO v11 model, the YOLO v11-EMA model achieves increases of 1.7, 0.4, and 0.7 percentage points in Precision, Recall, and mAP, respectively. Relative to the YOLO v11-MSDA model, the YOLO v11-EMA model exhibits a slightly lower Recall, while its Precision and mAP are improved by 0.4 and 1.2 percentage points, respectively. For the YOLO v11-iRMB model, although its mAP is comparable to that of the YOLO v11-EMA model, its Precision and Recall remain lower. Furthermore, compared with the YOLO v11-SCSA, YOLO v11-LSKA, YOLO v11-CGA, and YOLO v11-SEAM models, the YOLO v11-EMA model demonstrates superior performance in all three metrics. Collectively, these results indicate that the YOLO v11-EMA model exhibits the best overall performance in grape disease detection tasks.

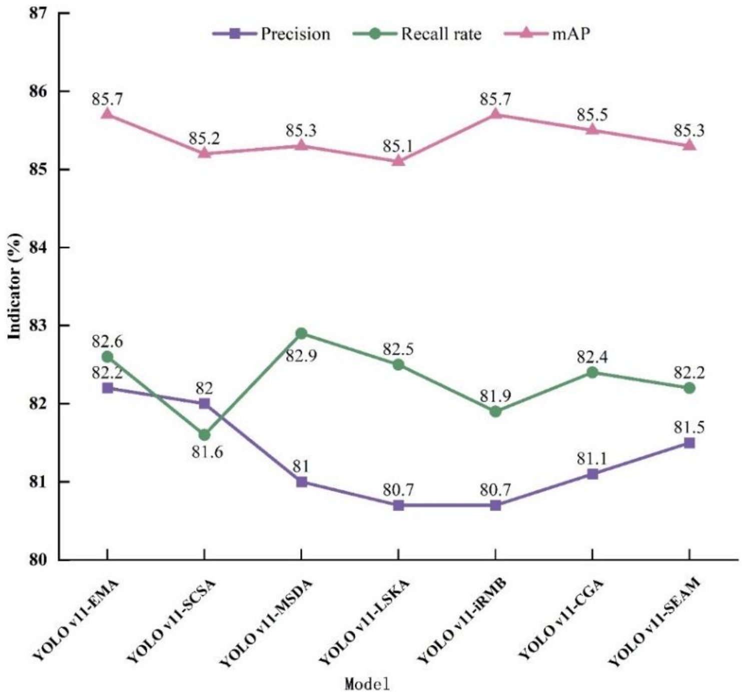

Fig. 8 - Comparative results of different attention modules

Ablation Experiment

The effectiveness of the improved PWE-YOLO model was evaluated through ablation experiments conducted on the custom-built grape disease dataset, with YOLO v11 serving as the baseline model. These experiments were designed to assess the impact of each individual module on overall model performance. The results of the ablation study are presented in Table 4, where "√" indicates the inclusion of a module and "×" indicates its exclusion. Here, P represents the C3k2-PConv module, W corresponds to the WaveletPool module, and E denotes the EMA attention mechanism.

Table 4

Results of ablation experiment

Model	WaveletPool	EMA	C3k2-PConv	P/%	R/%	mAP/%
YOLO v11	×	×	×	80.5	82.2	85.0
YOLO v11-P	×	×	√	81.5	82.7	85.7
YOLO v11-W	√	×	×	83.2	80.0	85.5
YOLO v11-E	×	√	×	82.2	82.6	85.7
YOLO v11-PW	√	×	√	83.9	83.2	85.8
YOLO v11-PE	×	√	√	84.5	83.4	86.6
YOLO v11-WE	√	√	×	83.3	83.1	86.4
PWE-YOLO	√	√	√	85.0	83.6	86.9

As presented in Table 4, the introduction of the C3k2-PConv module into the YOLO v11 model results in the YOLO v11-P model. Compared with the baseline YOLO v11 model, YOLO v11-P achieves improvements of 1.0, 0.5, and 0.7 percentage points in Precision, Recall, and mAP, respectively, while its number of parameters and floating-point operations are reduced by 8.13% and 6.35%, respectively. The integration of the WaveletPool module into YOLO v11 produces the YOLO v11-W model. Compared with YOLO v11, YOLO v11-W improves Precision and mAP by 2.7 and 0.5 percentage points, respectively, while its number of parameters and floating-point operations are reduced by 15.89% and 14.29%, respectively. Similarly, the addition of the EMA attention mechanism results in the YOLO v11-E model. Compared with the baseline model, YOLO v11-E achieves improvements of 1.7, 0.4, and 0.7 percentage points in Precision, Recall, and mAP, respectively, while its number of parameters and floating-point operations slightly increase. These results indicate that the individual incorporation of the C3k2-PConv, WaveletPool, and EMA modules can effectively enhance the detection performance of the model.

The simultaneous integration of the C3k2-PConv and WaveletPool modules into the YOLO v11 model produces the YOLO v11-PW model. Compared with the YOLO v11-P and YOLO v11-W models, YOLO v11-PW achieves improvements in Precision, Recall, and mAP, while its number of parameters and floating-point operations is significantly reduced relative to YOLO v11, YOLO v11-P, and YOLO v11-W. The combination of the C3k2-PConv module and the EMA attention mechanism results in the YOLO v11-PE model. YOLO v11-PE demonstrates enhanced Precision, Recall, and mAP compared with YOLO v11-P and YOLO v11-E, while its number of parameters and floating-point operations are reduced compared with YOLO v11 and YOLO v11-E, with a slight increase in floating-point operations relative to YOLO v11-P. Similarly, the integration of the WaveletPool module and the EMA attention mechanism yields the YOLO v11-WE model. Compared with YOLO v11-W and YOLO v11-E, YOLO v11-WE achieves improvements in Precision, Recall, and mAP, while its number of parameters and floating-point operations is significantly lower than in YOLO v11 and YOLO v11-E, with a slight increase in floating-point operations relative to YOLO v11-W.

The simultaneous integration of the C3k2-PConv module, the WaveletPool module, and the EMA attention mechanism into the YOLO v11 model produces the PWE-YOLO model. Compared with the baseline YOLO v11 model, PWE-YOLO achieves improvements of 4.5, 1.4, and 1.9 percentage points in Precision, Recall, and mAP, respectively, while reducing the number of parameters and floating-point operations by 24.03% and 10.05%, respectively. PWE-YOLO outperforms all other improved models across the evaluated metrics, demonstrating superior detection performance. The results of the ablation experiments confirm the effectiveness and rationality of the proposed design. By integrating the C3k2-PConv, WaveletPool, and EMA modules, the model's detection accuracy is effectively enhanced, while the computational cost is substantially reduced. These improvements make the proposed method particularly suitable for target detection tasks under constrained computational resources, highlighting its clear advantage for grape disease detection in controlled environments.

Comparative Experiment of Different Models

To further validate the advancement of the PWE-YOLO model proposed in this paper, a comparative experiment was conducted between the PWE-YOLO model and the YOLO v8 *(Ren et al, 2024)*, YOLO v9 *(Kalezhi and Shumba, 2025)*, YOLO v10 *(Zheng et al, 2025)*, YOLO v11, YOLO v12 *(Zhao et al, 2025)*, and YOLO v13 models. In the experimental design, all models were trained and tested under the same hardware and software environment to eliminate the influence of external environmental factors on the experimental results. The grape disease dataset constructed in this study was used, and reasonable hyperparameters were set to ensure the reliability and accuracy of the experimental results. The results are summarized in Table 5.

Table 5

Performance comparison of different models

Model	P/%	R/%	mAP/%	Parameters	FLOPs
YOLO v8	82.8	77.6	85.0	3.01×10^6	8.1×10^9
YOLO v9	79.7	80.8	85.1	2.62×10^6	10.7×10^9
YOLO v10	81.7	78.4	84.1	2.70×10^6	8.2×10^9
YOLO v11	80.5	82.2	85.0	2.58×10^6	6.3×10^9
YOLO v12	83.8	74.8	81.9	2.51×10^6	5.8×10^9
YOLO v13	83.5	79.8	84.6	2.45×10^6	6.2×10^9
PWE-YOLO	85.0	83.6	86.9	1.96×10^6	5.1×10^9

As shown in Table 5, the PWE-YOLO model achieves a Precision of 85.0%, Recall of 83.6%, and mAP of 86.9%, with a parameter count of 1.96 × 10^6 and floating-point operations of 5.1 × 10^9. Compared with YOLO v8, YOLO v9, YOLO v10, YOLO v11, YOLO v12, and YOLO v13, the Precision of PWE-YOLO increases by 2.2, 5.3, 3.3, 4.5, 2.2, and 2.5 percentage points, respectively; Recall increases by 6.0, 2.8, 5.2, 1.4, 8.8, and 3.8 percentage points, respectively; and mAP improves by 1.9, 1.8, 2.8, 1.9, 5.0, and 2.3 percentage points, respectively. In terms of computational complexity, PWE-YOLO demonstrates significant advantages over the compared models. The number of parameters is reduced by 34.88%, 25.19%, 27.40%, 24.03%, 21.91%, and 20.00%, respectively, while floating-point operations are decreased by 37.03%, 52.34%, 37.80%, 19.05%, 12.07%, and 17.74%, respectively. These results indicate that PWE-YOLO not only achieves higher accuracy in grape disease detection but also reduces computational complexity, featuring a lightweight architecture and faster inference speed. Overall, the proposed model demonstrates superior performance across Precision, Recall, and mAP, underscoring its effectiveness and suitability for real-time target detection in resource-constrained agricultural environments.

Analysis of Detection Results for Different Diseases

This study focuses on the recognition of multiple grape diseases, including downy mildew, powdery mildew, gray mold, canker, and Grapevine Leafroll-associated Virus (GLRaV) disease. Given the distinct manifestations of these diseases, the performance of the proposed PWE-YOLO model was evaluated separately for each disease type. To verify the recognition accuracy, comparative analyses were conducted between the original model and the improved PWE-YOLO model for all five disease categories.

Table 6

Comparison of model recognition accuracy for five diseases before and after improvement

Model	mAP (%)	Precision				
		downy mildew (%)	powdery mildew (%)	gray mold (%)	canker (%)	GLRaV (%)
YOLO v11	85.0	79.6	81.1	77.9	83.9	80.0
PWE-YOLO	86.9	84.8	85.5	82.2	87.0	85.5

Table 6 presents the recognition results for five grape diseases obtained using both the baseline YOLO v11 model and the improved PWE-YOLO model. The introduction of the C3k2-PConv module allows the model to perform convolution operations exclusively on valid regions during disease detection, effectively mitigating the impact of feature loss or damaged areas caused by insufficient lighting or unfavorable shooting angles. This approach prevents interference from missing regions in the final detection outcomes. The incorporation of the WaveletPool module enables effective suppression of downsampling aliasing while maintaining a lightweight network structure, significantly improving the retention of high-frequency features such as edges, patterns, and color variations in disease regions. Furthermore, the addition of the EMA attention mechanism enhances the model's focus on small targets, thereby increasing the detection accuracy of small lesions. As shown in Table 6, the integration of these three improvements leads to increases of 5.2, 4.4, 4.3, 3.1, and 5.5 percentage points in detection accuracy for the five grape diseases, respectively, compared with the baseline YOLO v11 model.

As shown in Table 6, through the three proposed improvements, the detection accuracy of the five grape diseases in the PWE-YOLO model increased by 5.2, 4.4, 4.3, 3.1, and 5.5 percentage points, respectively, compared with the original YOLO v11 model. These results indicate that the proposed PWE-YOLO model demonstrates superior performance across all disease types.

Analysis of Detection Results for Different Diseases

To comprehensively evaluate the detection performance of the proposed PWE-YOLO model, both PWE-YOLO and the baseline YOLO v11 models were applied to images from the test set. The visual detection results produced by the two models were compared, as illustrated in Figure 9.

The detection results are presented in Figure 9. For grape disease images with few targets, such as gray mold and Grapevine Leafroll-associated Virus (GLRaV) disease, the PWE-YOLO model accurately detects the disease regions, whereas the YOLO v11 model, although capable of detecting both diseases, exhibits slightly lower accuracy. For images containing multiple disease targets, such as downy mildew, powdery mildew, and canker, detection is challenged by overlapping symptoms and subtle disease manifestations, resulting in higher rates of missed and false detections.

Nevertheless, the PWE-YOLO model successfully detects a larger number of disease instances while maintaining lower rates of missed and false detections, whereas the YOLO v11 model exhibits missed detections for downy mildew and canker, and false detections for powdery mildew. These visualization results further validate the superiority and practical applicability of the PWE-YOLO model in grape disease detection tasks.

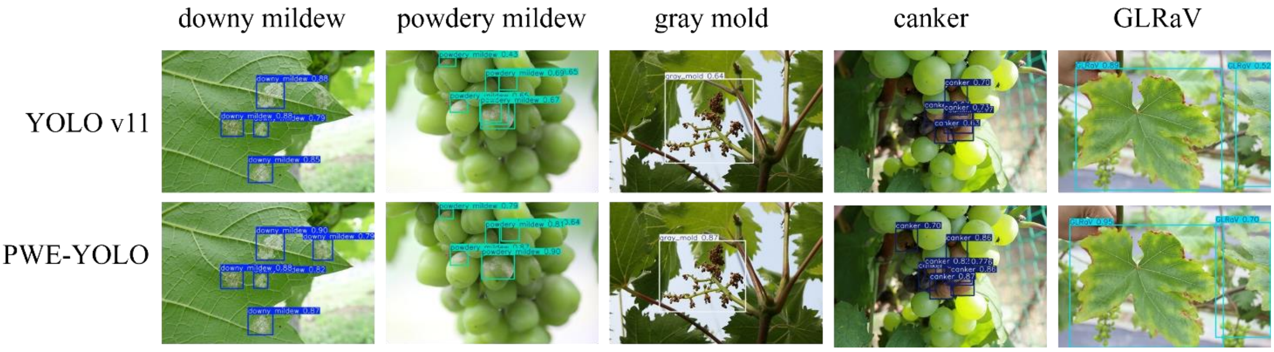

Fig. 9 - Detection results of different models

Practical Application in Vineyard Disease Inspection

To further demonstrate the practical applicability of the proposed method, a vineyard inspection robot platform was employed for field image acquisition. The platform consists of a mobile inspection robot, a robotic arm, and a depth camera mounted at the end of the robotic arm. During field operation, the robot moves along the vineyard rows, while the robotic arm adjusts the position and viewing angle of the camera to capture grape leaf images under natural environmental conditions. The acquired images are then input into the improved PWE-YOLO model for disease detection. The model outputs the disease category, confidence score, and location of the diseased region. Therefore, the integration of the inspection robot platform and the proposed detection model provides a feasible technical solution for intelligent grape disease monitoring, early warning, and precision vineyard management.

(a) (b) (c)

Fig. 10 - Field application of the vineyard inspection robot for grape disease detection

(a) Overall view of the mobile inspection robot in vineyard rows;
(b) image acquisition using the robotic arm and depth camera;
(c) close-up view of the depth camera mounted at the end of the robotic arm.

CONCLUSIONS

To achieve accurate detection of grape diseases in controlled environments, a PWE-YOLO model was proposed based on YOLO v11. By integrating the C3k2-PConv module, WaveletPool module, and EMA attention mechanism, the PWE-YOLO model improves detection accuracy, reduces floating-point operations and parameter count, and enhances inference speed.

The PWE-YOLO model achieves a Precision of 85.0%, Recall of 83.6%, and mAP of 86.9%, with a parameter count of 1.96×10^6 and floating-point operations of 5.1×10^9. Compared with the original YOLO v11 model, Precision, Recall, and mAP are improved by 4.5, 1.4, and 1.9 percentage points, respectively, while the number of parameters and floating-point operations are reduced by 24.03% and 10.05%, respectively. Furthermore, when compared with YOLO v8, YOLO v9, YOLO v10, YOLO v11, YOLO v12, and YOLO v13, PWE-YOLO demonstrates clear advantages across multiple key metrics, including Precision, Recall, and mAP. In practical detection scenarios, PWE-YOLO effectively reduces missed detections, demonstrating its potential for real-time, accurate, and efficient grape disease monitoring in facility environments.

ACKNOWLEDGEMENT

We would like to acknowledge the financial support from Innovation Team Fund for Fruit Industry of Modern Agricultural Technology System in Shandong Province (SDAIT-06-12), the Key R&D Program of Shandong Province, China (2024TZXD038,2024TZXD045), and Yantai Major Scientific and Technological Innovation Project (2024ZDCX025).

REFERENCES

[1] 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.

[2] Chen, J., Kao, S. H., He, H., Zhuo, W., Wen, S., Lee, C. H., & Chan, S. H. G. (2023). Run, don't walk: chasing higher FLOPS for faster neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12021-12031).

[3] Fu, Z., Yin, L., Cui, C., & Wang, Y. (2024). A lightweight MHDI-DETR model for detecting grape leaf diseases. *Frontiers in plant science, 15*, 1499911.

[4] Jiang, Y., Fang, Y., & Deng, L. (2025). PDCNet: A lightweight and efficient robotic grasp detection framework via Partial Convolution and knowledge distillation. *Computer Vision and Image Understanding, 259*, 104441.

[5] Jin, T., Shen, Z., & Geng, H. (2025). Optimized YOLOv11m for real-time high-speed railway catenary defect detection. *Scientific Reports*.

[6] Kalezhi, J., & Shumba, L. (2025). Cassava crop disease prediction and localization using object detection. *Crop Protection, 187,* 107001.

[7] Liu, B., Ding, Z., Tian, L., He, D., Li, S., & Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Frontiers in plant science, 11,* 1082.

[8] Li, X., Hou, H., Yu, Y., Fan, Z., & Zhao, X. (2025). YOLO-LitePV: a lightweight detection algorithm for photovoltaic panel defects. *Nondestructive Testing and Evaluation*, 1-28.

[9] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., & Huang, Z. (2023, June). Efficient multi-scale attention module with cross-spatial learning. In *ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1-5). IEEE.

[10] Poblete-Echeverría, C., Hernández, I., Iñiguez, R., Gutiérrez, S., Barrio, I., & Tardáguila, J. (2025). Using Artificial intelligence for automatic and fast detection of downy mildew symptoms in grapevine canopies. *European Journal of Agronomy, 170,* 127755.

[11] Qu, Y., & Zhang, F. (2025). FWL-YOLO: a lightweight model for barcode detection and recognition in express delivery waybills. *Signal, Image and Video Processing, 19*(3), 261.

[12] Ren, J., Zhang, H., Wang, G., Dai, C., Teng, F., & Li, M. (2024). Real-time grape disease detection model based on improved YOLOv8s. *INMATEH-Agricultural Engineering, 72*(1).

[13] Rahman, M. U., Liu, X., Wang, X., & Fan, B. (2024). Grapevine gray mold disease: Infection, defense and management. *Horticulture Research, 11*(9), uhae182.

[14] Raj, A., Dawale, M., Wayal, S., Khandagale, K., Bhangare, I., Banerjee, S., ... & Gawande, S. (2025). YOLO-ODD: an improved YOLOv8s model for onion foliar disease detection. *Frontiers in Plant Science, 16*, 1551794.

[15] Sun, J., Li, Y., Zhao, Q., Guo, Z., Li, N., Hai, T., ... & Chen, D. (2022). Cascade wavelet transform based convolutional neural networks with application to image classification. *Neurocomputing, 514*, 285-295.

[16] Williams, T., & Li, R. (2018, February). Wavelet pooling for convolutional neural networks. In *International conference on learning representations*.

[17] Xie, Y., Du, D., & Liu, Y. (2025). RE-YOLO: Vehicle Detector Based on Knowledge Distillation YOLOv8 for Autonomous Driving Scenarios. *Journal of Computing in Civil Engineering, 39*(6), 04025084.

[18] Xu, W., & Wang, R. (2023). ALAD-YOLO: An lightweight and accurate detector for apple leaf diseases. *Frontiers in plant science, 14*, 1204569.

[19] Yang, Y., Jiao, L., Liu, X., Li, L., Liu, F., Yang, S., & Zhang, X. (2024). Efficient LWPooling: Rethinking the wavelet pooling for scene parsing. *IEEE Transactions on Circuits and Systems for Video Technology, 34*(9), 8481-8493.

[20] Zhen, Y., Liu, P., Jin, Y., & Yang, G. (2025). Grape leaf disease detection and identification based on improved YOLOv8. *Transactions of the Chinese Society of Agricultural Engineering, 41*(14), 148-154.

[21] Zhang, N., Zhang, E., Qi, G., Li, F., & Lv, C. (2025). Lightweight grape leaf disease recognition method based on transformer framework. *Scientific Reports, 15*(1), 28974.

[22] Zhang, L., & Zeng, Q. (2024). Grape disease detection method based on styleGAN2-ADA and improved YOLOv7. *Trans. Chin. Soc Agric. Machinery., 55*, 241-252.

[23] Zhang, Z., Feng, B., & Liu, H. (2026). TS-YOLO: a small traffic sign detection algorithm for various harsh driving conditions in bad weather. *Scientific Reports*.

[24] Zheng, B., Angkawisittpan, N., Huang, L., & Sonasang, S. (2025). RSP-YOLOv11n multi-module optimized algorithm for insulator defect detection in UAV images. *Scientific Reports, 15*(1), 35426(13)

[25] Zhao, M., Cui, B., Yu, Y., Zhang, X., Xu, J., Shi, F., & Zhao, L. (2025). Intelligent Detection of Tomato Ripening in Natural Environments Using YOLO-DGS. *Sensors, 25*(9), 2664.

[26] Zhang, X., Zhanfeng, H. O. U., & Chuanzhong, X. (2022). Design and experiment of recognition system for coated red clover seeds based on machine vision. *INMATEH-Agricultural Engineering, 66*(1), pp.62-72. DOI: https://doi.org/10.35633/inmateh-66-06