



OPEN A lightweight improved YOLOv8 method for intelligent detection of pine wilt disease

Gang Chen¹, Mengquan Wu^{1,6}✉, Longxing Liu², Shikuan Wang³, Xiang Wang⁴, Xuqing Zhang⁵, Jinyi Lv¹, Yijie Tang¹, Bowen Liu¹, Hongyu Kan¹, Yuying Ran¹ & Huining Lai¹

Pine wood nematode disease (PWD) is one of the most devastating forest diseases worldwide, often described as the “cancer” of pine trees due to its rapid and large-scale lethality. Early and accurate detection of infected trees is essential for interrupting the transmission cycle and mitigating the risk of further spread. However, current monitoring methods suffer from limited efficiency and insufficient precision. To address these challenges, this study introduces PWD-YOLO-D, an intelligent detection model for PWD based on unmanned aerial vehicle (UAV) remote sensing imagery and the YOLOv8 deep learning framework. The proposed model integrates an Efficient Multi-scale Cross-Attention (EMCA) mechanism to enhance feature representation across multiple scales and heterogeneous backgrounds; incorporates a Self-Ensemble Attention Module (SEAM) as the detection head to improve robustness in identifying occluded and overlapping diseased crowns; and adopts the Focaler-IoU loss function to refine localization accuracy and improve discrimination of complex samples. Experimental results indicate that the improved PWD-YOLO-D model outperforms the original YOLOv8 by 4.0% points in AP@0.5 and 7.3% points in AP@0.5:0.95, while reducing the Parameters by 0.48 MB. These enhancements provide strong technical support and data-driven evidence for the timely detection and precise management of infected pine trees.

Keywords Pine wilt disease, UAV remote sensing, Object detection, YOLOv8, Deep learning, Forest health monitoring

Pine wilt disease (PWD), caused by the pine wood nematode (*Bursaphelenchus xylophilus*), is one of the most devastating diseases affecting pine forests worldwide¹. Once introduced, the disease can lead to rapid and large-scale mortality of host pine trees within a short time frame². Often referred to as the “cancer” of pine trees, PWD poses a severe threat to forest ecosystems³ and ranks among the most destructive forest diseases globally⁴. Since its first invasion into China in 1982, the pine wood nematode has killed billions of pine trees, causing direct economic losses and ecosystem service value losses amounting to tens of billions of yuan. This has placed China’s approximately 60 million hectares of pine forests under significant threat⁵. As of now, the disease has spread to 663 counties and districts across 18 provinces⁶ (municipalities), including Jilin Province, presenting a highly complex and urgent challenge for prevention and control efforts⁷.

Currently, the primary strategy for controlling PWD outbreaks involves the timely identification of discolored pine trees during the outbreak phase⁸, aiming to interrupt the transmission pathway and curb further disease spread⁹. The main monitoring approaches include ground-based surveys, satellite remote sensing, and unmanned aerial vehicle (UAV) monitoring¹⁰. However, due to the vast extent of pine forests, complex topography, and uneven distribution of infected trees, traditional manual inspection methods face several limitations, including high costs, intensive labor, extended time requirements, and the risk of omission. These challenges hinder a comprehensive understanding of the spatial distribution of infected trees in mountainous regions, potentially causing missed windows for effective control and leading to rapid disease escalation¹¹. Satellite remote sensing offers the advantage of large-scale coverage and macro-level monitoring capabilities¹², with benefits such as low cost, high efficiency, and broad operational scope compared to ground-based surveys. Nevertheless, its

¹College of Resources and Environmental Engineering, Ludong University, Yantai 264039, China. ²International Institute for Earth System Science, Nanjing University, Nanjing 210023, China. ³School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China. ⁴Yantai Municipal Forestry Bureau, Yantai 264000, China. ⁵Yantai City Fushan District Forestry Resource Monitoring and Management Service Center, Yantai 265599, China. ⁶Shandong HaoYu Space Information Technology Co., Ltd., Yantai 264039, China. ✉email: ld_wmq@ldu.edu.cn

effectiveness is limited by factors such as spatial resolution, atmospheric interference, and satellite revisit cycles, making it difficult to accurately locate discolored pine trees within critical time frames.

UAVs have emerged as a novel remote sensing platform for forest monitoring, offering high spatial resolution, rapid deployment capabilities, and low operational costs¹³. With the ability to respond in near real time and cover large forested areas efficiently, UAVs provide more precise data for monitoring and have been increasingly applied in the detection of diseased trees¹⁴. For instance, Syifa et al.¹⁵ utilized artificial neural networks (ANN) and support vector machines (SVM) to distinguish PWD-infected trees from healthy ones using UAV-based remote sensing data. Rao et al.¹⁶ applied an NDVI-based threshold segmentation algorithm to drone imagery to extract potential regions of interest (ROI) for PWD detection. Iordache et al.¹⁷ employed airborne multispectral and hyperspectral data and used the Random Forest (RF) algorithm to compare the classification accuracy of these datasets for PWD monitoring. While traditional machine learning methods have shown promise in PWD detection, they largely depend on hand-crafted features, which are labor-intensive to design and often inadequate in capturing the complex spatial and spectral characteristics of forest environments captured by UAVs. Moreover, the dynamic nature of forest ecosystems imposes additional challenges to model robustness and accuracy, thereby limiting the generalization capability of these methods across diverse forest types and geographic regions.

Deep learning technologies have shown remarkable capability in capturing the color and texture features of infected trees. Among them, the YOLO (You Only Look Once) series of object detection models¹⁸ are particularly notable for their high detection accuracy and real-time performance. These models significantly outperform traditional machine learning methods by enhancing shallow feature extraction and improving recognition accuracy, making them highly promising for the rapid identification of PWD and for facilitating timely disposal of infected trees¹⁹. For example, Wu et al.²⁰ applied YOLOv3 and Faster R-CNN to detect PWD and achieved an accuracy of 0.739 on their dataset. Chen et al.²¹ proposed a YOLOv5-based detection framework that integrates a Vision Transformer and CNN, achieving an accuracy of 90.04% in rapid PWD monitoring. Despite their promising performance, YOLO models still encounter challenges in complex environments, including false positives and false negatives caused by occlusion and dense object distribution, which can degrade overall detection reliability. Furthermore, traditional convolutional operations involve a large number of parameters and high computational costs. Therefore, exploring novel model architectures and optimization strategies is essential for achieving a better balance between accuracy, efficiency, and generalizability.

To address these challenges, this study proposes a lightweight object detection model named PWD-YOLO-D, built upon the YOLOv8 framework and designed for UAV-based deployment to enable precise and efficient monitoring of PWD. The main contributions of this study are as follows:

- (1) To address the limitations in capturing fine-grained visual features of infected foliage, this study introduces a novel Efficient Multi-scale Cross-Attention (EMCA) mechanism²², which enhances the model's ability to extract subtle color and texture cues by facilitating cross-channel and cross-scale feature interaction and fusion. This significantly improves sensitivity to pathological differences in leaves and boosts the precision of disease detection.
- (2) To address the lack of robustness in complex forest environments, a Self-Ensemble Attention Module (SEAM)²³ is incorporated as the detection head, leveraging multi-view feature fusion and consistency regularization to enhance detection performance under conditions such as occlusion, dense canopy overlap, and background clutter.
- (3) To address the limitations of traditional bounding box regression loss, the original CIoU loss is replaced with Focaler-IoU²⁴. This improved loss function incorporates dynamic loss weighting, monotonicity constraints, and advanced IoU optimization, which not only maintains high detection accuracy but also significantly enhances localization precision.

Materials and methods

Study area

The Kunyu Mountain Nature Reserve, located in Muping District, Yantai City, Shandong Province, lies in the eastern part of the Shandong Peninsula (121°37'00"–121°51'00"E, 37°12'20"–37°18'50"N) and is the only national-level nature reserve in Shandong classified as a forest ecosystem (Fig. 1). It serves as the native habitat and primary natural distribution area of Chinese red pine (*Pinus tabulaeformis*), and is renowned for containing the world's most well-preserved red pine forests. As such, it is often referred to as the "Plant Kingdom of Jiaodong." The pine forest ecosystem within the reserve remains largely undisturbed, with red pine forests primarily forming pure, continuous stands. In recent years, Kunyu Mountain has become a high-incidence area for PWD, with the epidemic spreading from neighboring regions and posing severe challenges to prevention and control efforts. This makes it a representative and ideal location for studying the transmission dynamics and control strategies of PWD.

UAV data acquisition and processing

This study conducted UAV flight operations between September 1 and October 1, 2023, utilizing the HY-300 vertical takeoff and landing fixed-wing aerial survey drone, which provided a ground sampling resolution of 0.03–0.07 m.

The acquired raw orthophotos were segmented into smaller image patches of 640 × 640 pixels. Through manual screening, a total of 2155 images containing suspected PWD-infected pine trees were selected. The dataset was then split into training and validation subsets at an 80:20 ratio, comprising 1724 and 431 infected sample images, respectively. Image annotation was performed using the open-source software Labelme, with annotation data saved in .txt format to facilitate subsequent model training. The key UAV flight parameters

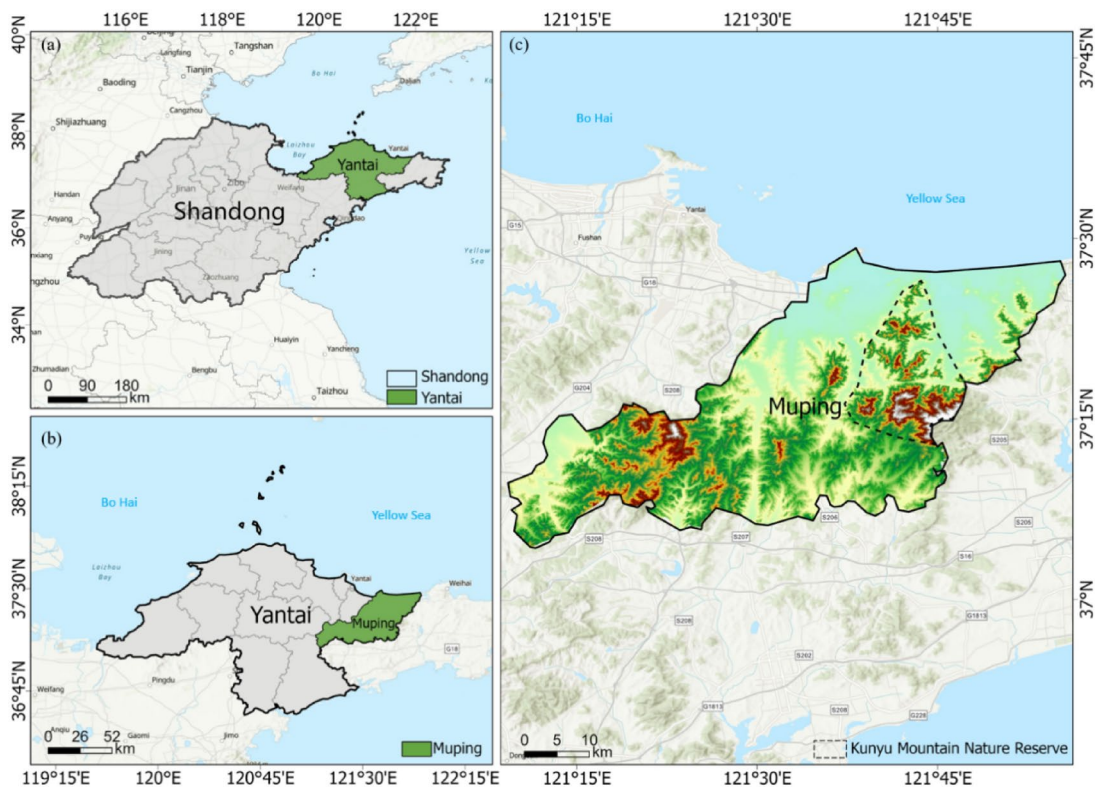


Fig. 1. Location of the study area: (a) Yantai City in Shandong Province; (b) Muping District in Yantai; (c) Kunyu Mountain Nature Reserve in Muping District. The map was created by the authors using ArcGIS 10.8 (<https://www.esri.com>) based on the 2024 provincial, municipal, and county-level administrative boundary dataset of China (Approval Number: GS(2024)0650).

Parameter	Value
Flight altitude (m)	150
Flight speed (m/s)	15
Forward overlap (%)	80
Side overlap (%)	80
Image capture interval (s)	8

Table 1. UAV flight parameters.

employed in this study are summarized in Table 1. Multi-scale evidence of pine wilt disease is illustrated in Fig. 2, including (a) UAV remote sensing imagery, (b) ground photographs of infected pine trees, and (c) microscopic observations of *Bursaphelenchus xylophilus*.

Overview of PWD-YOLO-D

YOLOv8 (You Only Look Once version 8) is a state-of-the-art real-time object detection model developed by Ultralytics²⁵. Its architecture comprises three main components: Backbone, Neck, and Head, delivering a robust combination of real-time inference speed, high detection accuracy, and ease of deployment. Within the YOLOv8 family, the YOLOv8s model strikes an optimal balance between accuracy and computational efficiency. Compared to YOLOv8n, YOLOv8s provides enhanced detection capability, while being more lightweight and efficient than the larger YOLOv8m, YOLOv8l, and YOLOv8x variants, making it particularly well-suited for real-time detection of PWD on UAV platforms.

Considering the complexities posed by rugged terrain, overlapping tree canopies, and substantial background interference in forested environments, this study proposes a lightweight, enhanced PWD detection model—PWD-YOLO-D—based on the YOLOv8s framework. The model architecture has been redesigned to address the specific challenges of PWD detection. The network structure is illustrated in the accompanying figure.

The model has been systematically enhanced across three key aspects: feature extraction, architectural design, and loss function optimization (Fig. 3). First, the EMCA mechanism has been integrated into the YOLOv8 backbone network. This attention module dynamically adjusts the response weights of critical feature channels,

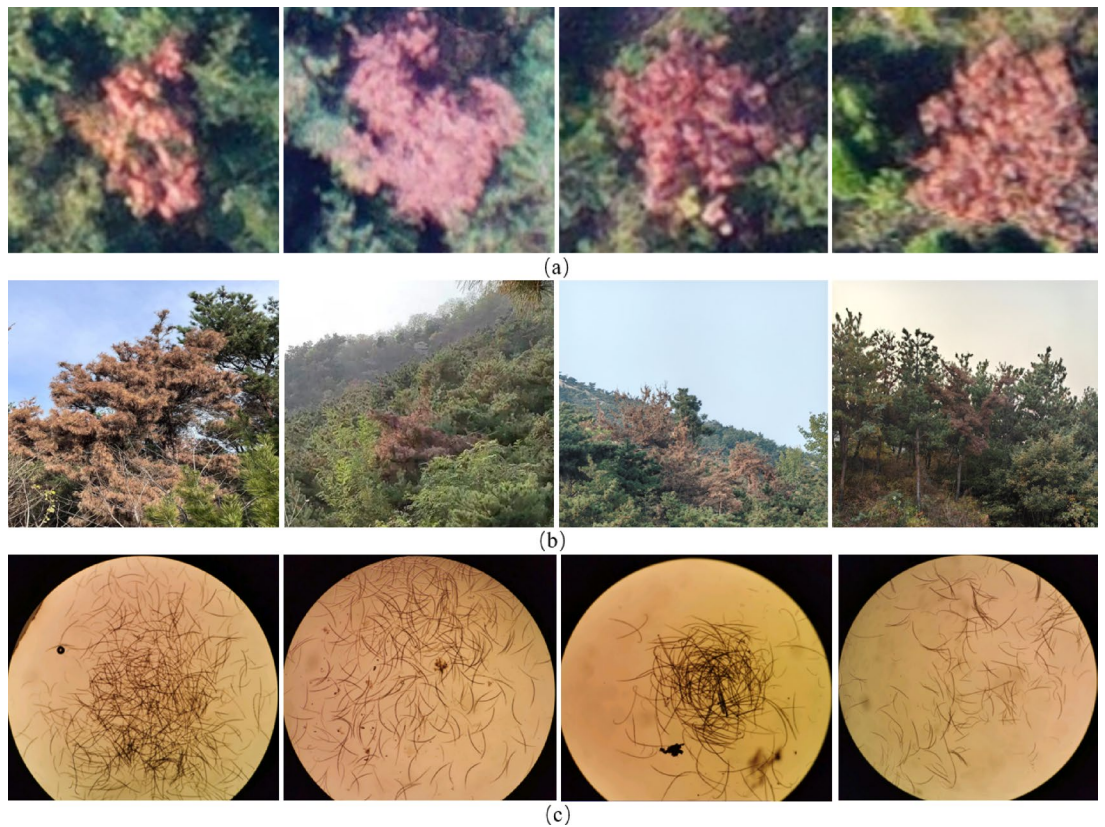


Fig. 2. Multi-scale evidence of pine wilt disease: **(a)** UAV remote sensing imagery acquired by the authors, **(b)** ground photographs of infected pine trees, and **(c)** microscopic observations of *Bursaphelenchus xylophilus*. All photographs including UAV imagery and ground images, were taken by the authors G.C., X.W., and B.L.

effectively suppressing redundant background information and enhancing the model's capability to represent features in multi-scale target scenarios.

Second, the conventional detection head has been replaced with the SEAM, which strengthens semantic feature extraction and fusion, thereby improving both the flexibility and accuracy of bounding box localization.

Finally, the localization regression loss function has been upgraded from the traditional CIoU loss to Focaler-IoU. This novel loss function optimizes regression accuracy and enhances robustness, particularly for challenging cases involving occluded or blurred target boundaries.

EMCA: efficient Multi-scale cross attention

In remote sensing imagery, the leaf discoloration caused by PWD often manifests as small spatial regions with subtle color variations and indistinct structural features. These weak disease signals are easily confounded by complex backgrounds such as healthy vegetation, forest floor, or shadow noise. To enhance the model's sensitivity to such subtle pathological cues, this study incorporates the EMCA mechanism into the YOLOv8 backbone network.

EMCA²⁶ is a lightweight attention module that integrates multi-scale feature modeling with cross-channel interaction mechanisms, enabling effective enhancement of feature representation while maintaining computational efficiency. The EMCA module performs parallel convolutional operations on input feature maps at multiple scales to capture feature responses under varying receptive fields, thereby improving the model's ability to detect targets at different spatial resolutions. By introducing a cross-channel attention mechanism, EMCA establishes interactive dependencies among feature channels across scales, overcoming the limitation of traditional attention mechanisms that only model within a single scale and channel. This mechanism employs a compact one-dimensional convolutional kernel to characterize channel-wise dependencies at each scale and dynamically adjusts the relative importance of features across scales via a cross-scale fusion strategy. Consequently, it effectively highlights critical disease-affected regions while suppressing background interference (Fig. 4).

Let the input feature map be denoted as $X \in R^C \times H \times W$. Multi-scale features are extracted by applying n convolution operations with different receptive fields (e.g., different kernel sizes or dilation rates):

$$F_s = \text{Conv}_s(X), \quad s \in \{1, 2, \dots, n\} \quad (1)$$

where Conv_s denotes the convolution operation at the s -th scale, and the output feature map is $F_s \in R^{C \times H_s \times W_s}$.

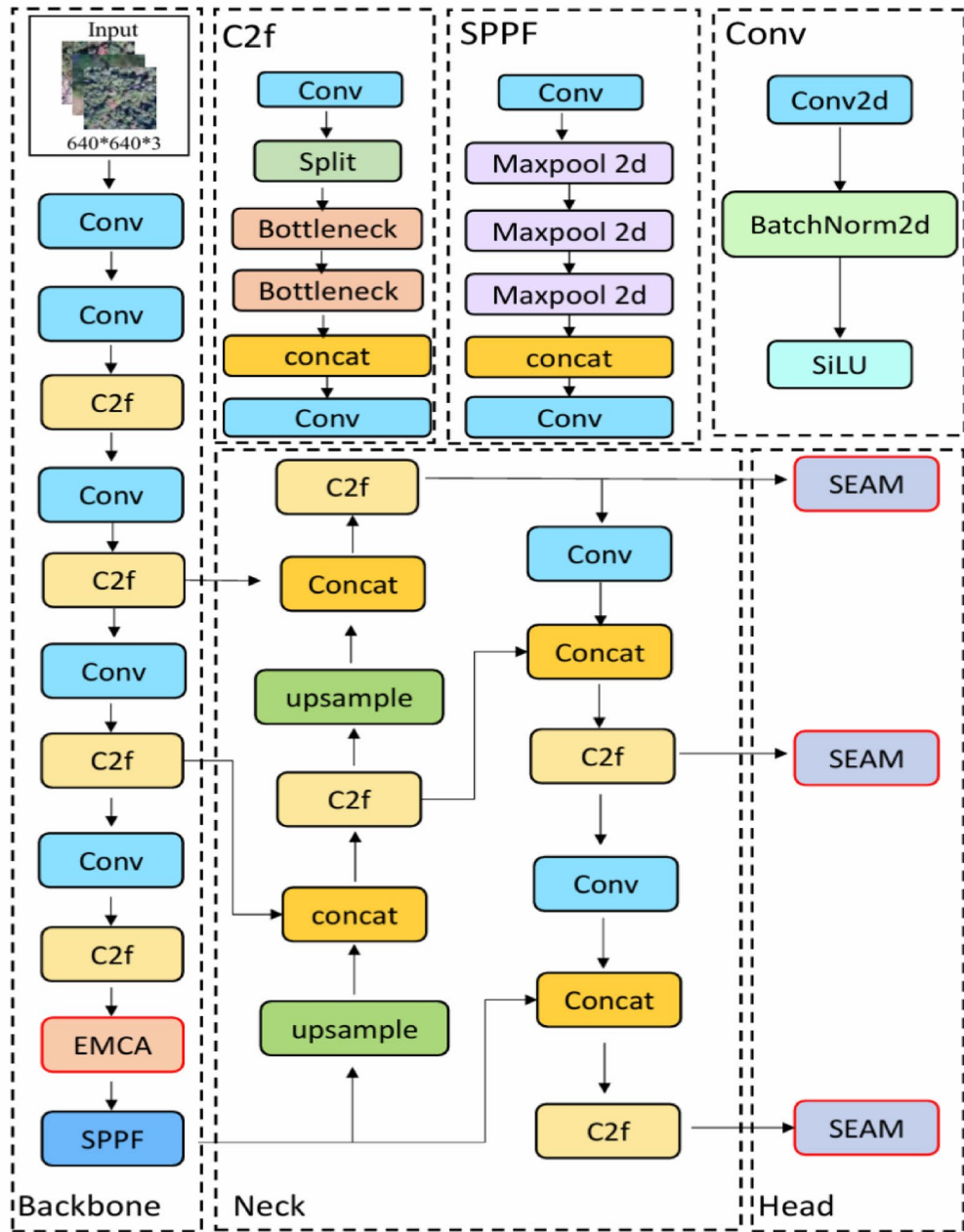


Fig. 3. Architecture of the proposed PWD-YOLO-D model.

Global average pooling (GAP) is then applied to each scale-specific feature map to obtain its corresponding channel descriptor vector:

$$Z_s = \frac{1}{H_s \times W_s} \sum_{i=1}^{H_s} \sum_{j=1}^{W_s} X_s(:, i, j) \tag{2}$$

This vector is fed into a one-dimensional convolution layer Conv1D_k to model inter-channel contextual dependencies. The channel attention weights are then obtained through a Sigmoid activation function:

$$a_s = \sigma(\text{Conv1D}_k(Z_s)) \tag{3}$$

Here, *k* represents the kernel size of the 1D convolution, which is adaptively set according to the number of channels.

The attention vector is applied to the original feature map via element-wise multiplication to achieve attention-based weighting:

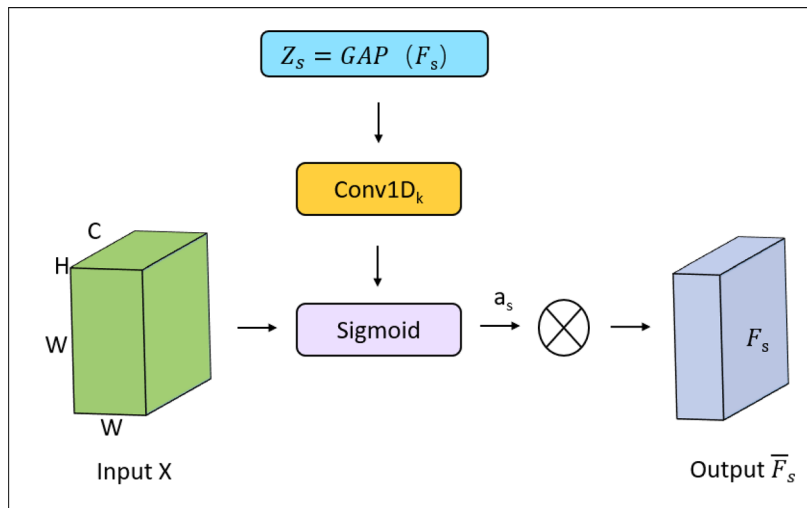


Fig. 4. Schematic of the EMCA attention mechanism. F_s denotes the feature map at the s -th scale, Z_s is the channel descriptor obtained via GAP, a_s is the channel attention weight produced by Conv1D_k and Sigmoid, and \bar{F}_s is the enhanced feature map.

$$\bar{F}_s = a_s \odot F_s \tag{4}$$

The attention-enhanced features from all scales are subsequently fused to produce the final output:

$$F^{EMCA} = \text{Fusion}(\bar{F}_1, \bar{F}_2, \dots, \bar{F}_n) \tag{5}$$

where Fusion(·) denotes the feature fusion operation, such as concatenation followed by channel compression or weighted summation.

SEAM: Self-Ensembling attention module detection head

Traditional detection heads often face challenges in effectively modeling the contextual semantic relationships between targets, especially in complex natural scenes characterized by dense small targets, severe occlusions, and substantial background interference. The SEAM module²⁷ employs a dual-branch architecture that jointly models spatial details and high-level semantic features, thereby enhancing the model’s discriminative capability. The semantic enhancement branch utilizes multiple convolutional blocks with dilated convolutions to expand the receptive field without sacrificing spatial resolution, enabling the capture of richer contextual information. This facilitates improved perception of long-range dependencies and macro-level lesion characteristics, such as large-scale reddening or crown degradation. Concurrently, the spatial preservation branch maintains the original spatial details of the input features via shallow convolutional layers, focusing on extracting fine-grained details like edges and textures, which is particularly beneficial for detecting localized and subtle lesion areas. The outputs from both branches are fused and further processed through channel and spatial attention mechanisms to amplify responses in key regions while suppressing redundant background noise, thus optimizing the detector’s performance in both classification and localization tasks (Fig. 5).

The process can be expressed as:

$$F_s = \text{Conv}_{\text{dilated}}(F), \quad F_p = \text{Conv}_{\text{shallow}}(F) \tag{6}$$

$$F_{\text{cat}} = \text{Concat}(F_s, F_p) \tag{7}$$

$$F_{\text{att}} = \text{CA}(F_{\text{cat}}) \cdot \text{SA}(F_{\text{cat}}) \tag{8}$$

$$F_{\text{out}} = \text{Detect}(F_{\text{att}}) \tag{9}$$

Here, F_s denotes the feature map extracted from the semantic path, and F_p represents the feature map extracted from the spatial path. CA(·) and SA(·) denotes the final output of the object detection network. F_{out} denotes the final output of the object detection network.

Focaler-IoU loss function

To overcome the challenges in accurately localizing difficult samples, this study redesigns the regression loss function of YOLOv8. Building upon the traditional CIoU loss, the Focaler-IoU loss function²⁸ is incorporated to form a novel Focaler-CIoU loss. This loss function improves upon CIoU through two key enhancements: firstly, it increases the model’s focus on hard-to-localize samples, enabling more effective attention to targets embedded in complex backgrounds or with high ambiguity, thereby boosting localization accuracy; secondly, it introduces a dynamic penalty mechanism that modulates the penalty strength according to the sample’s difficulty level,

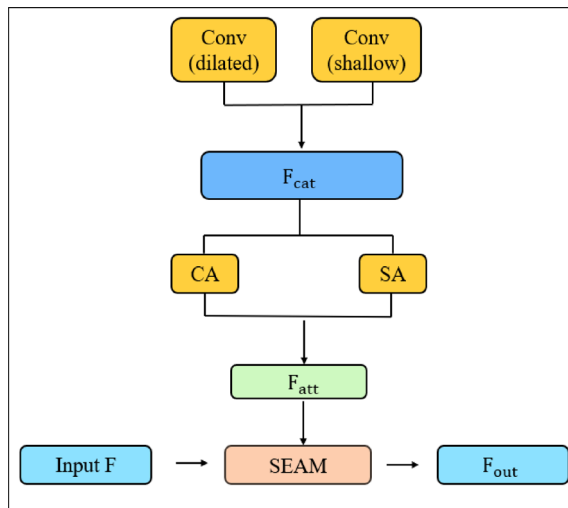


Fig. 5. Schematic diagram of the SEAM detection head.

imposing stronger constraints on low-IoU samples and further refining localization precision. The formulation of the Focaler-IoU loss is defined as follows:

$$\text{IoU}^{\text{focaler}} = \begin{cases} 0, & \text{IoU} < d \\ \frac{\text{IoU}-d}{u-d}, & d < \text{IoU} < u \\ 1, & \text{IoU} > u \end{cases} \quad (10)$$

In this definition, $\text{IoU}^{\text{focaler}}$ represents the re-weighted IoU, where IoU is the original intersection-over-union value, and $d, u \in [0, 1]$ are hyperparameters that control the sensitivity to sample difficulty. By adjusting d and u , the function can be tuned to emphasize different categories of regression samples. The corresponding loss is given by:

$$L_{\text{Focaler-IoU}} = 1 - \text{IoU}^{\text{focaler}} \quad (11)$$

When incorporated into the CIoU-based regression loss, the resulting Focaler-CIoU is defined as:

$$L_{\text{Focaler-IoU}} = L_{\text{CIoU}} + \text{IoU} - \text{IoU}^{\text{Focaler}} \quad (12)$$

This formulation effectively combines the geometric awareness of CIoU with the adaptive focusing capability of Focaler-IoU, thereby improving the model's localization performance on challenging samples.

Evaluation metrics

Given the high contagion risk of PWD, timely and accurate detection of infected trees is critical for effective disease prevention and control. This study comprehensively evaluates detection performance using key metrics including Precision (P), Recall (R), and Average Precision (AP).

Precision quantifies the proportion of true positive predictions among all positive predictions made by the model, representing the ratio of correctly identified positive samples to the total predicted positives. Recall⁴ measures the proportion of correctly detected positive samples relative to the total number of actual positive samples. AP is computed as the area under the Precision-Recall (P-R) curve for each category, reflecting the model's overall ability to balance precision and recall across varying confidence thresholds. This metric provides a thorough assessment of the model's detection performance within each category.

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

$$AP = \int_0^1 P \cdot RdR \quad (15)$$

Experimental environment

The computational platform employed in this study is equipped with an Intel® Core™ i9-9900 K processor and an NVIDIA A800 80GB PCIe GPU, running CUDA version 12.4 to ensure ample GPU memory resources for efficient training and inference of large-scale remote sensing imagery. The software environment utilizes Python

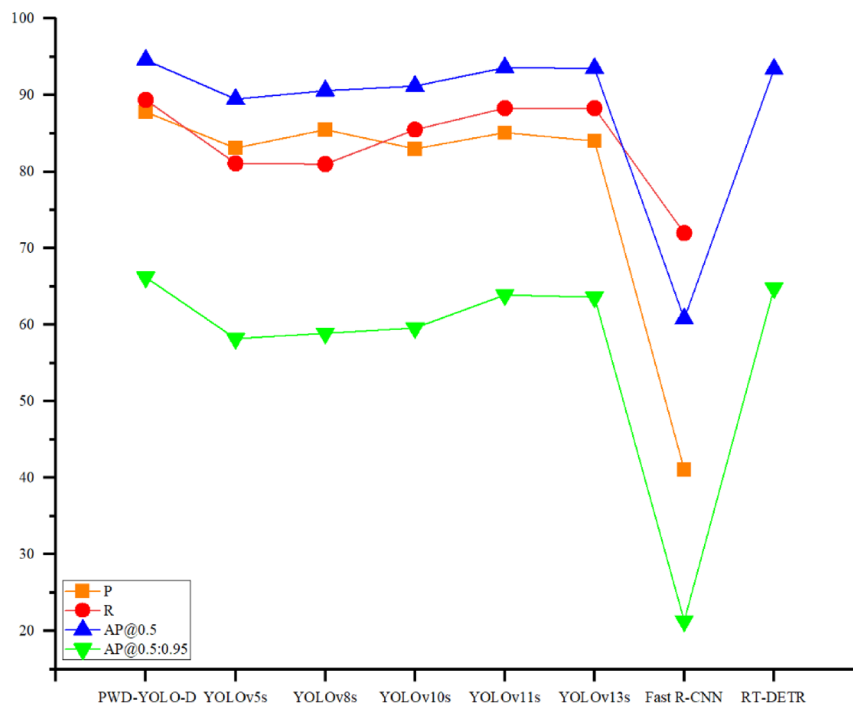


Fig. 6. Performance comparison of different models.

Model	P (%)	R (%)	AP@0.5 (%)	AP@0.5:0.95 (%)	Parameters (MB)
PWD-YOLO-D	87.8	89.4	94.6	66.2	10.65
YOLOv5s	83.1	81.1	89.5	58.2	7.01
YOLOv8s	85.5	81	90.6	58.9	11.13
YOLOv10s	83	85.5	91.2	59.6	7.21
YOLOv11s	85.1	88.3	93.6	63.9	9.46
YOLOv13s	84.0	88.3	93.5	63.6	9.00
RT-DETR	-	-	93.4	64.8	652
Fast R-CNN	41.04	71.98	60.79	21.3	137.10

Table 2. Performance comparison of different Models. Precision (P) and Recall (R) are not provided for RT-DETR as the original implementation does not include these metrics. Significant values are in bold.

3.9.21 and PyTorch 2.5.1, with the object detection framework based on Ultralytics YOLOv8 version 8.2.50, ensuring compatibility with the latest features and modules. The training process consists of 100 epochs to guarantee full model convergence. A batch size of 32 is used per iteration to optimize computational efficiency and memory utilization. The initial learning rate is set to 0.01 to promote stable and effective convergence. All models uniformly process input images resized to 640×640 pixels.

Results

Ablation study and comparative analysis

To validate the effectiveness and generalization capability of the proposed PWD-YOLO-D model, a series of comparative experiments were conducted on a self-constructed pine wilt disease dataset. The performance of PWD-YOLO-D was evaluated against several representative models, including YOLOv5s, YOLOv10s, YOLOv11s, YOLOv13s, DETR-based models, and the two-stage detector Fast R-CNN.

Experimental results, as illustrated in Fig. 6 and summarized in Table 2 demonstrate that the improved PWD-YOLO-D model exhibits significant advantages across all evaluation metrics. Specifically, it achieves a precision of 87.8%, recall of 89.4%, AP@0.5 of 94.6%, and AP@0.5:0.95 of 66.2%, outperforming other models in both detection accuracy and robustness.

Compared with other YOLO series models, PWD-YOLO-D exhibits clear advantages in detection performance. Relative to YOLOv5s, which achieves a precision of 83.1%, recall of 81.1%, AP at 0.5 of 89.5%, AP at 0.5 to 0.95 of 58.2%, and a parameter size of 7.01 megabytes, PWD-YOLO-D improves AP at 0.5 by 5.1 points and AP at 0.5 to 0.95 by 8.0 points. In comparison with YOLOv8s, with a precision of 85.5%, recall of 81%, AP at 0.5 of 90.6%, AP at 0.5 to 0.95 of 58.9%, and a parameter size of 11.13 megabytes, PWD-YOLO-D increases

YOLOv8s						
EMA	SEAM	Focaler-IoU	AP@0.5 (%)	AP@0.5:0.95 (%)	Parameters (MB)	GFLOPs
			90.6	58.9	11.13	28.4
√			93.4	63.4	11.13	28.4
√	√		93.8	64.1	10.65	25.7
√	√	√	94.6	66.2	10.65	25.7

Table 3. Ablation study results after adding different modules to the baseline YOLOv8s.

Loss function	P (%)	R (%)	AP@0.5 (%)	AP@0.5:0.95 (%)
Focal-IoU	87.8	89.4	94.6	66.2
CloU	84.7	87.6	93.2	62.7
MDPIoU	85.4	87.4	93.7	63.5
InnerIoU	87.6	84.1	92.3	61.9
WISEIoU	83.3	86.6	92.2	60.9

Table 4. Comparison of different loss functions.

AP at 0.5 by 4.0 points and AP at 0.5 to 0.95 by 7.3 points. YOLOv10s, although benefiting from a smaller and simpler architecture, achieves a precision of 83%, recall of 85.5%, AP at 0.5 of 91.2%, and AP at 0.5 to 0.95 of 59.6%, which still limits its practical detection capability.

Lightweight variants YOLO11s and YOLO13s also show competitive performance. YOLO11s attains a precision of 85.1%, recall of 88.3%, AP at 0.5 of 93.6%, AP at 0.5 to 0.95 of 63.9%, and a parameter size of 9.46 megabytes. YOLO13s achieves a precision of 84%, recall of 88.3%, AP at 0.5 of 93.5%, AP at 0.5 to 0.95 of 63.6%, and a parameter size of 9.00 megabytes. Despite surpassing YOLOv5s, YOLOv8s, and YOLOv10s overall, both models remain inferior to PWD-YOLO-D in detecting small targets under complex backgrounds, highlighting the effectiveness of the proposed structural improvements and module integration strategies.

Among non-YOLO models, RT-DETR achieves an AP at 0.5 of 93.4% and AP at 0.5 to 0.95 of 64.8%, but its large parameter size of 652 megabytes limits inference efficiency. The two-stage detector Fast R-CNN achieves a precision of 41.04%, recall of 71.98%, AP at 0.5 of 60.79%, and AP at 0.5 to 0.95 of 21.3%, markedly underperforming the YOLO series in both detection capability and practical deployment.

Contribution of different modules to model performance

To verify the effectiveness of each module in enhancing model performance, this study conducted systematic ablation experiments using a self-constructed dataset. The EMCA attention mechanism, SEAM detection head, and Focaler-IoU loss function were sequentially integrated into the baseline YOLOv8s model (Table 3).

Through systematic performance comparisons, it was found that introducing the EMCA mechanism led to improvements in AP@0.5 and AP@0.5:0.95 by 2.8 and 4.5% points, respectively, while maintaining the parameter count at 11.13 MB and GFLOPs at 28.4. This verifies that the EMCA mechanism enhances model accuracy without increasing computational complexity.

Following the integration of the SEAMHead detection head, AP@0.5 and AP@0.5:0.95 further improved by 0.5 and 0.7% points, respectively, while the number of parameters decreased to 10.65 MB and GFLOPs reduced to 25.7, indicating that this module not only enhances feature extraction but also contributes to model compression and efficiency.

Upon incorporating all components, the model achieved the best overall performance, with AP@0.5 reaching 94.6% and AP@0.5:0.95 rising to 66.2%, while maintaining the parameter count at 10.65 MB and GFLOPs at 25.7. Compared to the baseline, AP@0.5 and AP@0.5:0.95 increased by 4.0 and 7.3% points, respectively. These results demonstrate that the combined application of all modules synergistically maximizes the model's capability for small object detection while reducing computational cost.

Performance comparison of different loss functions

In object detection tasks, the accuracy of bounding box regression critically influences the detector's localization performance. As a fundamental component for optimizing bounding box predictions, the design rationality of the loss function plays a decisive role in overall model efficacy. To evaluate the specific impact of various IoU-based loss functions on detection performance, this study introduces five mainstream schemes—CloU, Focal-IoU, MDPIoU, InnerIoU, and WISEIoU—while keeping the model architecture and training configurations constant. Their performance is systematically compared using four key metrics: P, R, AP@0.5, and AP@0.5:0.95, as summarized in Table 4.

Experimental results comparing different IoU loss functions demonstrate that Focal-IoU achieves the best performance across all evaluation metrics (Fig. 7), with a precision of 87.8%, recall of 89.4%, AP@0.5 of 94.6%, and AP@0.5:0.95 of 66.2%, significantly enhancing detection effectiveness. Compared to the conventional CloU, Focal-IoU improves precision by 3.1% points, recall by 1.8% points, AP@0.5 by 1.4% points, and AP@0.5:0.95 by 3.5% points, indicating its superior fine-grained localization capability.

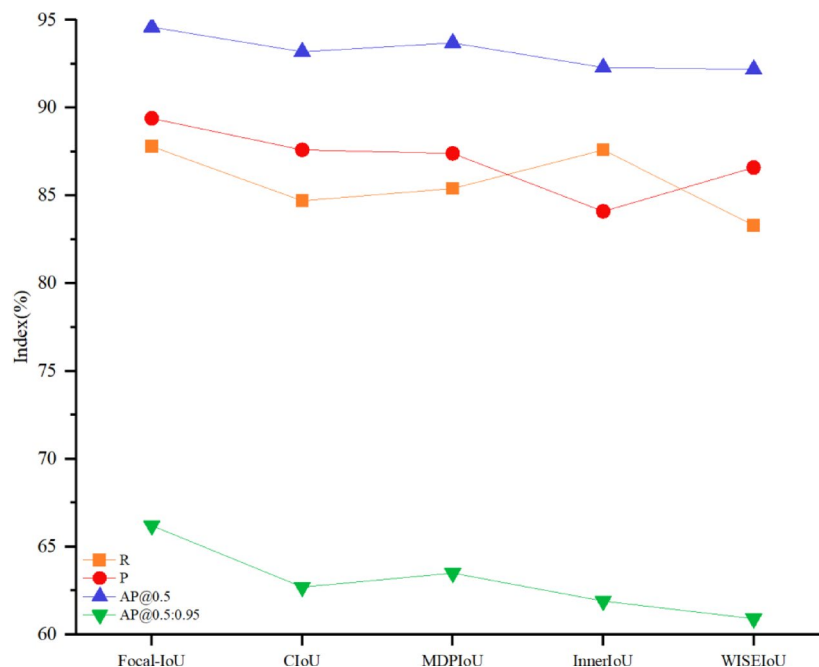


Fig. 7. Performance comparison of different loss functions.

MDPIoU exhibits stable performance with modest improvements over CIoU across all metrics, notably increasing AP@0.5:0.95 by 0.8% points, suggesting some optimization potential. InnerIoU achieves a precision of 87.6% but suffers a considerable drop in recall to 84.1%, reflecting its bias towards detection confidence at the expense of comprehensive target coverage. WISEIoU performs poorly across all metrics and is unsuitable for the current detection task.

In summary, Focal-IoU stands out as the most effective loss function in this study, significantly boosting detection accuracy and localization precision, and is thus recommended as the optimal choice.

Robustness evaluation of the model under complex backgrounds

To systematically evaluate the performance of PWD-YOLO-D in detecting PWD targets, this study selected five sets of images under varying backgrounds, lighting conditions, and occlusions. PWD-YOLO-D and other lightweight models were used to perform recognition and prediction on these images, and the results were exported for comparative analysis.

As illustrated in Fig. 8, under the bright background conditions of the first image set, all evaluated models demonstrated satisfactory detection performance. In the second image set, due to reduced image brightness and the presence of interfering objects, both YOLOv8s and YOLOv13s exhibited one false negative and one false positive, indicating susceptibility to environmental variations. In the fourth image set, both YOLOv8s and YOLOv13s produced a false positive, while in the fifth image set, characterized by numerous small targets and relatively blurred backgrounds, YOLOv8s failed to detect some small-scale diseased trees.

In contrast, PWD-YOLO-D, benefiting from its enhanced feature extraction and multi-scale detection capabilities, consistently achieved superior detection accuracy across all image sets. The model accurately identified small-scale diseased trees even under challenging conditions, such as complex backgrounds, low lighting, and object occlusion. No false negatives or false positives were observed in the test images, demonstrating its robust generalization ability and strong adaptability to diverse environmental conditions. These results highlight PWD-YOLO-D's reliability for practical applications in pine wilt disease monitoring across heterogeneous forest scenes.

Discussion

PWD is a severe forest pathology that poses significant threats to ecosystem stability and the forestry economy. Due to the small spatial scale of affected areas, heavy occlusion, and strong background interference in remote sensing images, traditional detection methods often suffer from limited accuracy, high false positive rates, and poor generalization capabilities in real-world forest environments.

To enhance the model's capability in detecting small objects and improve robustness in complex scenarios, previous studies have proposed various optimization strategies. For instance, Wang et al.²⁹ introduced a Global Attention Module (GAM) to strengthen the extraction of small object features, while Sun et al.³⁰ developed a Receptive Field and Direction-Induced Attention Network (RDIAN) to mitigate the inter-class imbalance between targets and backgrounds. During feature extraction, multi-receptive field convolutional structures were fused to capture multi-scale local features, significantly boosting the detection performance of small targets in infrared imagery. Although these methods have advanced small target representation, they often incur

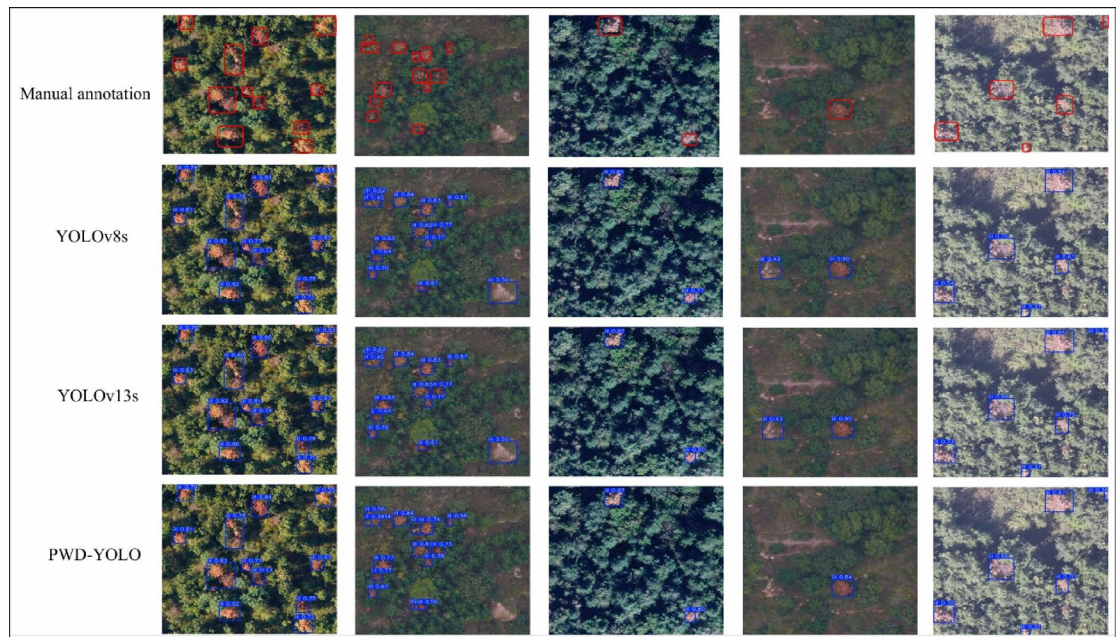


Fig. 8. Detection performance of different models under real complex backgrounds. The images were captured by the authors G.C., X.W., and B.L.

substantial computational overhead and increased parameter complexity, thereby limiting their feasibility for deployment on edge devices or in real-time applications.

This study addresses the challenges posed by dense small target distributions, subtle feature variations, and complex backgrounds in pine wood nematode disease detection through three key structural optimizations: the EMCA attention mechanism, the SEAM detection head, and the Focaler-IoU loss function. Each component targets distinct challenges and collectively enhances overall model performance.

The EMCA mechanism improves the model's sensitivity to fine-grained textures and crown lesions by concurrently extracting features across multiple scales, effectively mitigating the insufficient feature extraction of small targets in complex backgrounds commonly seen in traditional YOLO models. After integrating the EMCA module, the model's AP@0.5 increased by approximately 2.8% points and AP@0.5:0.95 by 4.5% points compared to the original YOLOv8, demonstrating significant improvements in scale adaptability and detailed texture representation. The SEAM detection head employs a dual-path design that combines spatial and semantic features, preserving spatial details while enhancing high-level semantic information. This facilitates better separation and identification of occluded targets and those with similar textures. Upon introducing the SEAM head, AP@0.5 and AP@0.5:0.95 further improved by approximately 0.4 and 0.7% points, respectively, while the model's parameter count decreased by about 0.48 MB. This module not only enhances feature representation efficiency but also effectively reduces redundant computations. The Focaler-IoU loss function significantly optimizes bounding box localization accuracy, particularly for challenging samples, by dynamically adjusting loss weights for low-IoU samples. This enhances regression precision for blurred boundaries and occluded objects. After replacing the original loss with Focaler-IoU, the model's AP@0.5 and AP@0.5:0.95 improved by 0.8 and 2.1% points, respectively, while precision and recall increased by 1.3 and 2.0% points, effectively reducing false negatives and false positives in complex forest environments.

Although the proposed model demonstrates strong robustness and accuracy in detecting small objects within complex backgrounds, several challenges remain to be addressed in practical applications. The model effectively identifies suspected pine wilt disease lesions across most forest environments; however, in regions characterized by high background noise, complex terrain, or severe canopy overlap, some weak or poorly defined lesion features may still be missed, particularly in cases of shallow reddening or atypical symptom presentations.

Currently, the model primarily relies on RGB imagery, which limits its ability to capture physiological abnormalities before the disease enters the latent stage, when significant color or texture changes have yet to manifest. Since the latent period is critical for epidemic prevention, future research should consider integrating multi-source remote sensing data, such as hyperspectral or thermal infrared imagery³¹, to detect subtle vegetation spectral changes and facilitate earlier disease detection during this stage.

Moreover, the model's performance varies under different lighting conditions. In non-ideal imaging environments—such as low light, backlighting, or high contrast scenarios—the contrast between disease features and background diminishes, resulting in reduced detection accuracy. This underscores the need for future work to improve the model's lighting invariance and stability under complex natural illumination³².

Furthermore, although key metrics such as mAP@0.5 and mAP@0.5:0.95 have significantly improved, there remains potential for optimizing inference computational cost. The current model size has been compressed to 10.65 MB, making it suitable for deployment on most edge devices; however, it still imposes considerable

computational demands when processing continuous imagery over extensive forest areas. Future efforts could explore techniques such as model architecture compression, parameter quantization, sparse training, and pruning to enhance computational efficiency while preserving detection accuracy.

Conclusion

This study addresses the challenges of insufficient accuracy and low efficiency encountered by existing detection methods in complex forest environments by proposing the PWD-YOLO-D detection model, which is optimized based on the YOLOv8s algorithm. By integrating the EMCA mechanism, the SEAM detection head, and the Focaler-IoU loss function, the model significantly enhances the accuracy and robustness of diseased tree identification, offering an innovative solution for intelligent forest disease prevention and control.

Experimental results demonstrate that the improved model achieves an AP@0.5 of 94.6% and an AP@0.5:0.95 of 66.2%, while reducing the total number of parameters by 0.48 megabytes. This represents a substantial performance improvement while maintaining a lightweight design, thus providing technical feasibility for deployment on UAV platforms and edge computing devices. Compared with commonly used object detection models such as Faster R-CNN, YOLOv5s, YOLOv8s, YOLOv10s, and YOLO13s, our model improves PWD detection accuracy by 33.81, 5.1, 4.0, 3.4, and 1.1% points in terms of AP@0.5, respectively.

Looking ahead, we plan to further optimize the PWD-YOLO-D model architecture and deployment strategies, integrating it with mobile platforms such as drones and ground robots to develop an integrated real-time monitoring and early warning system. This system will enable online detection in field environments and support dynamic adjustment of parameter configurations in response to environmental changes (e.g., seasonal variations and fluctuating light conditions), ensuring stable and robust detection performance. Ultimately, this will facilitate early identification and rapid response to pine wood nematode disease, effectively safeguarding forest ecosystem health and promoting sustainable forest management.

Data availability

The data are available from the corresponding author, Dr. Wu, upon request.

Received: 16 August 2025; Accepted: 16 October 2025

Published online: 20 November 2025

References

- Zhao, B. G., Futai, K., Sutherland, J. R. & Takeuchi, Y. *Pine Wilt Disease* (Springer, 2008).
- Brovkina, O., Cienciala, E., Surový, P. & Janata, P. Unmanned aerial vehicles (UAV) for assessment of qualitative classification of Norway spruce in temperate forest stands. *Geo-Spat Inf. Sci.* **21**, 12–20 (2018).
- Dixon, C. M., Robertson, K. M., Ulyshen, M. D. & Sikes, B. A. Pine savanna restoration on agricultural landscapes: The path back to native savanna ecosystem services. *Sci. Total Environ.* **818**, 151715 (2022).
- Wang, G. et al. A novel BH3DNet method for identifying pine wilt disease in Masson pine fusing UAS hyperspectral imagery and LiDAR data. *Int. J. Appl. Earth Obs Geoinf.* **134**, 104177 (2024).
- Liu, F. et al. Refined assessment of economic loss from pine wilt disease at the subcompartment scale. *Forests* **14**, 139 (2023).
- Ryss, A. Y., Kulinich, O. A. & Sutherland, J. R. Pine wilt disease: A short review of worldwide research. *Stud. China.* **13**, 132–138 (2011).
- Ye, J. R. Epidemic status of pine wilt disease in China and its prevention and control techniques and counter measures (2019).
- Zhou, X. et al. A monitoring method for pine wilt disease infested discolored and deceased pine trees removal information based on DDPNet network and Bi-temporal UAV imagery. *Remote Sens. Appl. : Soc. Environ.* **38**, 101530 (2025).
- Jin, W. T. et al. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *Proc. Natl. Acad. Sci. USA* **118**, e2022302118 (2021).
- Zhang, B. et al. A spatiotemporal change detection method for monitoring pine wilt disease in a complex landscape using high-resolution remote sensing imagery. *Remote Sens.* **13**, 2083 (2021).
- Pan, J., Lin, J. & Xie, T. Exploring the potential of UAV-based hyperspectral imagery on pine wilt disease detection: Influence of spatio-temporal scales. *Remote Sens.* **15**, 2281 (2023).
- Riley, J. R. *Remote sensing in entomology* (1989).
- Duarte, A., Borralho, N., Cabral, P. & Caetano, M. Recent advances in forest insect pests and diseases monitoring using UAV-based data: A systematic review. *Forests* **13**, 911 (2022).
- Xu, S. et al. Automatic pine wilt disease detection based on improved YOLOv8 UAV multispectral imagery. *Ecol. Inf.* **84**, 102846 (2024).
- Syifa, M., Park, S. J. & Lee, C. W. Detection of the pine wilt disease tree candidates for drone remote sensing using artificial intelligence techniques. *Eng* **6**, 919–926 (2020).
- Rao, D. et al. Deep learning combined with balance mixup for the detection of pine wilt disease using multispectral imagery. *Comput. Electron. Agric.* **208**, 107778 (2023).
- Iordache, M. D. et al. A machine learning approach to detecting pine wilt disease using airborne spectral imagery. *Remote Sens.* **12**, 2280 (2020).
- Wang, G. et al. Traffic sign detection method based on improved YOLOv8. *Sci. Rep.* **15**, 19385 (2025).
- Ye, W. et al. Pine pest detection using remote sensing satellite images combined with a multi-scale attention-UNet model. *Ecol. Inf.* **72**, 101906 (2022).
- Wu, B. et al. Application of conventional UAV-based high-throughput object detection to the early diagnosis of pine wilt disease by deep learning. *Ecol. Manag.* **486**, 118986 (2021).
- Chen, Y. et al. An efficient approach to monitoring pine wilt disease severity based on random sampling plots and UAV imagery. *Ecol. Indic.* **156**, 111215 (2023).
- Luo, Q. et al. An efficient multi-scale channel attention network for person re-identification. *Vis. Comput.* **40**, 3515–3527 (2024).
- Luo, M. & Ji, S. Cross-spatiotemporal land-cover classification from VHR remote sensing images with deep learning based domain adaptation. *ISPRS J. Photogramm. Remote Sens.* **191**, 105–128 (2022).
- Liang, M. et al. Research on detection of wheat tillers in natural environment based on YOLOv8-MRF. *Smart Agric. Technol.* **10**, 100720 (2025).
- Wang, S. et al. An advanced multi-source data fusion method utilizing deep learning techniques for fire detection. *Eng. Appl. Artif. Intell.* **142**, 109902 (2025).

26. Bakr, E. M., El-Sallab, A. & Rashwan, M. EMCA: Efficient multiscale channel attention module. *IEEE Access*. **10**, 103447–103461 (2022).
27. Xu, Y. et al. Self-ensembling attention networks: Addressing domain shift for semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence* Vol. 33, 5581–5588 (2019).
28. Zhang, H. & Zhang, S. Focaler-iou: More focused intersection over union loss. Preprint at <https://arxiv.org/abs/2401.10525> (2024).
29. Wang, S. et al. Detection of pine wilt disease using drone remote sensing imagery and improved YOLOv8 algorithm: A case study in Weihai, China. *Forests* **14**, 10 (2023).
30. Sun, H. et al. Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST. *IEEE Trans. Geosci. Remote Sens.* **61**, 1–13 (2023).
31. Deng, J. et al. RustQNet: Multimodal deep learning for quantitative inversion of wheat stripe rust disease index. *Comput. Electron. Agric.* **225**, 109245 (2024).
32. Lv, Z. & Pomeroy, J. W. Detecting intercepted snow on mountain needleleaf forest canopies using satellite remote sensing. *Remote Sens. Environ.* **231**, 111222 (2019).

Author contributions

M.W. conceived and supervised the study. G.C., M.W. and L.L. designed the research framework and methodology. S.W. and J.L. contributed to remote sensing data processing and annotation. G.C., X.W. and B.L. supported field data acquisition and validation. H.K. and Y.R. conducted model training and performance evaluation. H.L. contributed to visualization and result interpretation. G.C. wrote the main manuscript text. All authors contributed to manuscript revision and approved the final version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 42071385), the National Science and Technology Major Project of High-Resolution Earth Observation System (No. 79-Y50-G18-9001-22/23), the Research Topics of Yantai City Smart City Innovation Lab (No. 202310-04-ZHCS-01), and the Shandong Science and Technology SMEs Technology Innovation Capacity Enhancement Project (2022TSGC2317).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025