scientific reports

OPEN

YOLO-DP: A detection model of fifteen common rice diseases and pests

Tianxiang Zhou & Linjing Wei ✉

During rice cultivation, common rice diseases and pests such as Rice blast, Bacterial blight, Brown-planthopper and Leaf-folder will significantly affect the yield and quality. The current model is limited to detecting rice diseases or pests alone, and faces challenges such as the diversity of disease and pest traits, small detection targets, uneven light and complex background shading in paddy fields, resulting in low accuracy and adaptability of the model. In this study, a YOLO-DP (Diseases and Pests) model based on YOLOv8n model was proposed to detect fifteen common rice diseases and pests under complex conditions. First, the Triplet Attention mechanism is introduced into the network's Backbone to achieve cross-dimensional interaction between channels and spatial dimensions. Then, GLSA (the Global to Local Spatial Aggregation) module is used to improve the Neck of YOLOv8n, enhancing the effectiveness of feature representation. The WTConv (Wavelet Transform Convolution) is used to improve the C2f-BottleNeck of the original model, expanding the network's receptive field. Finally, the loss function is replaced with EIoU (Enhanced Intersection over Union) to reduce the position offset and shape mismatch of the predicted boxes. Experimental results demonstrate that the model achieves an average accuracy of 80.9%, a recall rate of 74.4%, a Mean Average Precision mAP50 of 77.8% and mAP95 of 50.1%, significantly outperforming the original YOLOv8n and mainstream detection models such as TOOD, Faster R-CNN and RT-DETR. This model exhibits exceptional performance in detecting rice diseases and pests in complex environments, providing robust technical support for rice growth monitoring and offering insights for the detection of other crop diseases.

Rice serves as the primary food source for more than half of the global population, making the prevention of rice diseases and pests essential for maintaining food security[1]. However, the emergence of rice diseases and pests poses a significant threat to crop yields and quality, leading to incalculable annual economic losses due to their impact on agriculture[2]. Common rice diseases can cause leaves to wither, affecting photosynthesis and damaging the stems, nodes, panicles, young branches, and grains[3]. In severe cases, these diseases can lead to total crop failure. Rice pests can damage the plant's vascular tissues, and large numbers of pests can cause the rice to fall over, leading to growth deformities and significantly impacting yield and quality[4].

To combat these threats more effectively, the development of accurate and efficient disease and pest detection methods has become a central focus of current research efforts[5]. Traditional monitoring methods for rice diseases and pests are largely based on eye observations by plant protection personnel, which has the following shortcomings: First, it is highly subjective. The identification results are significantly affected by individual factors such as the observer's experience and fatigue level, which can easily lead to misjudgment. Second, in large-scale planting scenarios, a large amount of manpower is required for field-by-field inspection, resulting in high monitoring costs. Third, it has poor timeliness. The outbreak of diseases and pests is often sudden, and manual inspection has a long cycle and slow response, which can easily miss the optimal window for prevention and control[6]. Traditional computer vision-assisted methods attempt to improve efficiency through image processing technologies (such as threshold segmentation and morphological feature extraction), they still have obvious limitations: highly dependent on manually designed features, and their robustness to complex field environments is poor. In the realm of target detection technology, deep learning algorithms offer innovative solutions for identifying rice diseases and pests, leveraging their prowess in feature extraction and pattern recognition[7]. First, the early identification and intervention of rice diseases and pests are crucial to safeguard rice

College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China. ✉email: wlj@gsau.edu.cn

growth and yield, mitigate yield losses due to diseases and pests, and enhance agricultural production efficiency[8]. Furthermore, the use of deep learning technology for the identification of diseases and pests can reduce the dependence on chemical pesticides, which is beneficial for environmental protection and the promotion of sustainable agriculture[9]. Finally, the application of deep learning technology strengthens the research basis of smart agricultural technology, promotes the automation and intelligence of agricultural production, and is expected to improve the level of food security on a global scale, and has important social value to promote social and economic development and improve the quality of people's life[10]. In recent years, rapid progress in computer hardware, particularly in processing speeds, along with advances in software technologies, has propelled the application of deep learning and image processing to the forefront of research in the domain of diseases and pests prevention and control[11].

Zhang et al.[12] have toke advantage of both CNN and transformer for accurate detection of leaf blast and brown spot diseases. The corresponding precision, recall, and F1-score were all over 0.96 with an AUC of up to 0.9987, and the corresponding loss rate was 0.0042. Uddin et al.[13] proposed a novel end-to-end training of convolutional neural network (CNN) and attention (E2ETCA) ensemble framework that fuses the features of two CNN-based state-of-the-art (SOTA) models along with those of an attention-based vision transformer model. Rajasekhar et al.[14] developed a Spider Monkey-based Random Forest (SMbRF) model for the precise detection and classification of rice leaf diseases, achieving a remarkable accuracy of 99.29%, sensitivity of 99.52%, precision of 98.76%, and specificity of 99%, demonstrating the model's scalability. Sangaiah et al.[15] have enriched the T-YOLOv4 network by integrating SPP (Spatial Pyramid Pooling)[16], CBAM (Convolutional Block Attention Module), SCFEM (Sand Clock Feature Extraction Module), Ghost modules, and additional convolutional layers, thereby enhancing the network's precision in identifying the three types of rice leaf diseases, the testing mean average precision (mAP) as 86%. Gao et al. have further enhanced the YOLOv5 by adding the CBAM (Convolutional Block Attention Module), optimizing the main branch gradient flow with the BottleNeck Block module of CSP[17] (Cross Stage Partial) in the Neck, and substituting the SPPF (Spatial Pyramid Pooling-Fast)[18] module with the S-SPPF module to detect the three kinds of rice leaf diseases, the detection result increased by 9.90% in terms of mAP. Song et al.[19] have introduced a novel YOLOv8-SCS architecture that incorporates SPD-Conv (Space-to-Depth Convolution)[20], the CG block (Context Guided block)[21], and Slide Loss, enabling the detection of 10 distinct rice pests, achieving an mAP of 87.9%, which is a 5.7% improvement over the original YOLOv8. Trinh et al.[22] have innovatively altered the loss function of the YOLOv8n model by incorporating Alpha-IoU and have implemented this enhanced model in IoT devices to detect the three most prevalent rice leaf diseases in Vietnam, the precision of the proposed model in this research has reached 89. 9% in the data set of 3175 image[23]. Table 1 shows the summary of AI models for agricultural diseases and pests detection in recent years. However, these target detection methods depend on the characteristic parameters of accurate extraction, feature extraction difficulty, weak anti-interference ability and poor universality. There are fewer types of diseases and pests detection, the detection accuracy and the applicability of these models still need to be improved.

This article first introduces materials and methods, the second step is to conduct experiments, give results and comparisons and the last step is to conclude. This model is on the basis of YOLOv8n, using Triplet Attention mechanism to model channel attention and spatial attention, using GLSA (Global to Local Spatial Aggregation) module to effectively fuse feature information at different levels, C2f-BottleNeck for improvement using WTConv (Wavelet Transform Convolution), replace the loss function with EIoU (Enhanced Intersection over Union) and propose a new model called YOLO-DP. This model can detect 15 different types of diseases and pests, a visualization system has also been created for the model to detect images and videos of diseases and pests, on the basis of the original YOLOv8n model, the accuracy rate increased by 2.9% to 77.8%, which can effectively detect more types of rice diseases and pests.

Materials and methods
Experimental environment configuration
During the experiment, in order to ensure the efficiency of model training, the cloud server AutodL was selected as the experimental platform, equipped with RTX 4090 (24 G) graphics card, the processor is 18 vCPU AMD EPYC 9754 128-Core Processor, the platform runs on the Ubuntu 20.04 LTS operating system, the deep learning framework is PyTorch 1.11.0, the development language is Python 3.8, and the computing platform is Cuda 11.3. During the training of the network model, the initial learning rate was set at 0.01, the number of iterations to 300 and the number of training batches to 64. The environment of the developed rice diseases and pests detection system is PyQt5 5.15.2, the development language is Python 3.9.20, framework is PyTorch 1.9.0.

Ref	Models	Data set	Accuracy
Zhang et al.[12]	CNN, Transformer	Leaf blast and Brown spot diseases	96%
Uddin et al.[13]	CNN, E2ETCA, SOTA	Mendeley, Kaggle Rice Diseases, Bangladesh Rice Research Institute	99.1%
Rajasekhar et al.[14]	SMbRF	Brown spot, Bacterial blight, Blast, and Tungro	99.29%
Sangaiah et al.[15]	T-YOLOv4, SPP, CBAM	Three types of rice leaf diseases	86%
Gao et al.[17]	YOLOv5, CBAM	Sheath blight, Rice blast, Bacterial blight	9.9% rise
Song et al.[19]	YOLOv8, SPD-Conv	Ten distinct rice pests	87.9%
Trinh et al.[22]	YOLOv8n, Alpha-IoU	Three most prevalent rice leaf diseases in Vietnam	89.9%

Table 1. Summary of AI models for agricultural diseases and pests detection.

Data set construction

The construction of high-quality data sets is the premise of model training. Existing public data sets do not collect rice diseases and rice pests at the same time, so this study constructed a new data set by collecting online and field photography, gathering data of fifteen common rice diseases and pests. The field photography in data set comes from Dongping Town, Chongming District, Shanghai, China, the camera used for image capture was the Apple iPhone 15, which collected nine types of rice diseases including: Rice blast, Bacterial blight, Brown spot, Rice dead heart, Rice downy mildew, Rice false smut, Rice sheath blight, Rice bacterial leaf streak and Rice tungro, as well as six types of rice pests including Brown-planthopper, Green-leafhopper, Leaf-folder, Rice-bug, Stem-borer and Whorl-maggot, these images are shown in Fig. 1. These data, combined with the public data set, cover the characteristics of pests and diseases in different regions and environments, fully ensure the diversity and practicability of the data set and can provide strong support for subsequent research and analysis.

The data set was divided into training, validation, and test groups in a 7:2:1 ratio. At the same time, use Baidu's intelligent data service Easy Data to annotate the missing labels of rice diseases and pests images, the annotation is the specific lesion area. To enhance the model's generalization and robustness, the original images were processed with transformations flipping and rotation, using four methods: brightening, dark, salt noise and Gaussian noise adjustments were made to the H (Hue), S (Saturation) and V (Value) components in the HSV color space to increase the diversity in color, at the same time to emulate different scenes of rice fields such as in sunny, rainy, morning, dusk and night. These methods and transformations are shown in Fig. 2, through these methods the data set was expanded to 20,129 images. Finally, the composition of the data set is presented in Table 2. Although the enhancement strategies have enriched data diversity to a certain extent, there are still potential limitations: the current environmental simulation mainly focuses on basic changes in light intensity and weather conditions, and has not yet fully covered the complex actual scenarios in the field.

YOLOv8n network structure improvement

YOLOv8n (You Only Look Once v8 Nano) is a real-time target detection model launched by Ultralytics, which provides frontier performance in terms of accuracy and speed. As a new iteration of the YOLO series, YOLOv8n composed of modules such as Input, Backbone, Neck, Head, Loss and Output, improve the performance of feature extraction and object detection[24].

The model YOLO-DP in this article will build on the YOLOv8n: Introducing Triplet Attention into the Backbone of the network captures cross-dimensional interactions between channels and spatial dimensions through three branching structures, thus improving the model's ability to understand and process features, in the mean time it has a negligible computational overhead. Improve YOLOv8n using GLSA modules to effectively fuse feature information at different levels, improve the context perception of the model, enhance the effectiveness of the feature representation. C2f-BottleNeck for improvement using WTConv by cascadading decomposition of wavelet transform can significantly expand the receptive field of the network without significantly increasing the number of parameters. Finally, the original loss function CIoU was replaced with EIoU by directly minimizing the width and height differences of the prediction box from the real box, considering the matching of position and size comprehensively. Reduce the position offset and shape mismatch of the prediction box, making the model converge faster. Based on these, we proposed a new model YOLO-DP, Fig. 3 is the structure diagram of YOLO-DP.
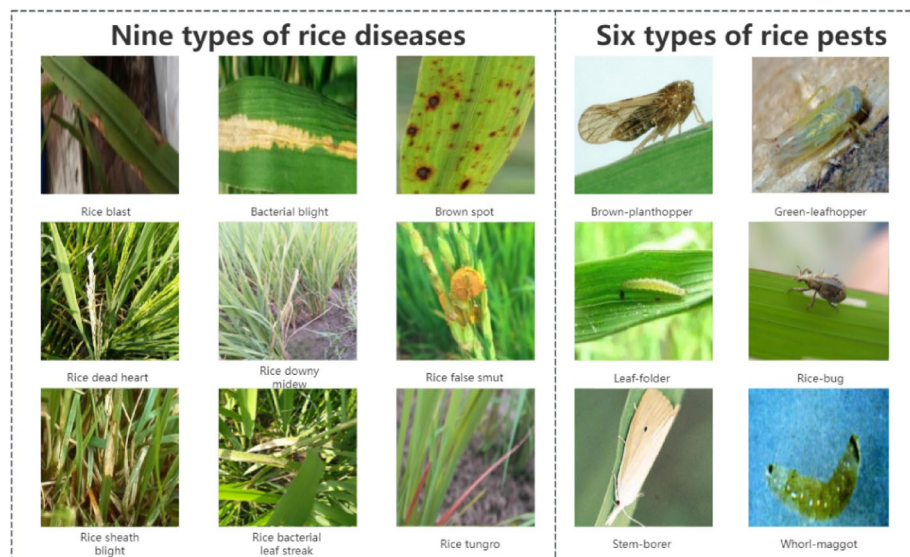

Fig. 1. Fifteen types of rice diseases and pests. This figure covering nine types of rice diseases including: Rice blast, Bacterial blight, Brown spot, Rice dead heart, Rice downy mildew, Rice false smut, Rice sheath blight, Rice bacterial leaf streak and Rice tungro, as well as six types of rice pests including Brown-planthopper, Green-leafhopper, Leaf-folder, Rice-bug, Stem-borer and Whorl-maggot.
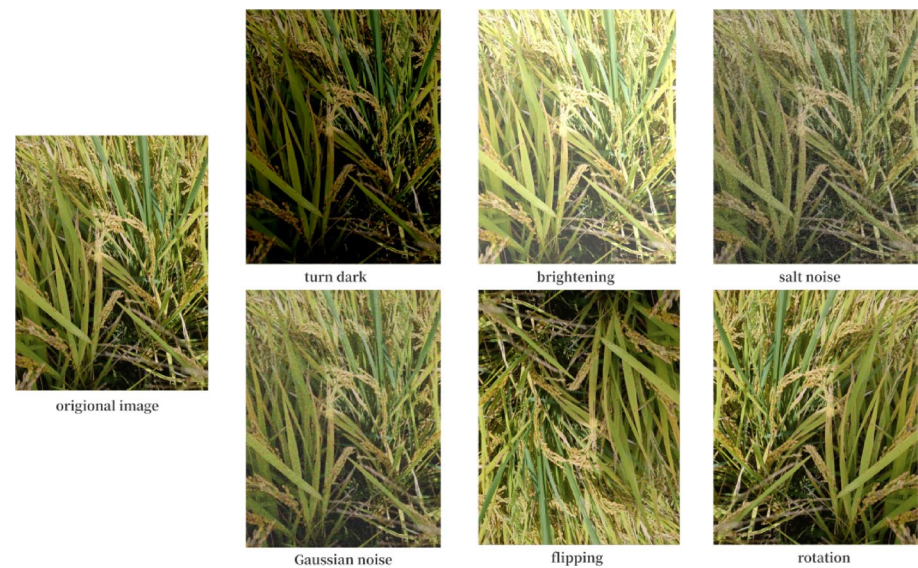
Fig. 2. Six types of data augmentation methods. This figure shows six types of data augmentation methods to emulate different scenes of rice fields such as in sunny, rainy, morning, dusk and night.

Name of diseases and pests	Training/images	Validation/images	Test/images
Brown-planthopper	867	213	117
Green-leafhopper	1050	264	117
Leaf-folder	1074	306	150
Rice-bug	882	225	132
Stem-borer	1014	312	153
Whorl-maggot	978	306	171
Rice false smut	858	256	135
Rice sheath blight	939	253	155
Rice bacterial leaf streak	1006	288	138
Rice tungro	732	209	108
Rice blast	977	293	155
Bacterial blight	909	248	128
Brown spot	954	279	142
Rice dead heart	989	286	140
Rice downy mildew	822	256	143

Table 2. Composition of the data set.

The main steps are collecting images of rice diseases and pests, adjusting them to a resolution of 640×640, augmenting the data set through image algorithms to enhance the model's generalization and robustness. Then training the preprocessed images and corresponding labels in the detection network to obtain the model weight files; using specific loss functions and optimizers during the training process to ensure optimal performance of the model for specific tasks. Finally, validating and testing the trained model weight files on test images to assess the model's performance and accuracy. Training metrics such as average accuracy, average recall rate, and mean precision are recorded to comprehensively evaluate the model's performance[25].

Triplet attention

In the complex rice field environment, the YOLOv8n model faces challenges with rice leaves, subtle diversity of rice ear disease characteristics, low contrast between infected and healthy leaves, and similar pest color to rice wheat ears, and similar leaf color. These factors make it difficult for the original model to accurately detect and locate the characteristics of rice pests and pests. To address these challenges, this study was used Triplet Attention, an innovative approach to attention mechanisms, leveraging a three-pronged architecture to discern cross-dimensional interplay within input data and calculate attention weights. This technique facilitates the creation of interdependencies among input channels or spatial positions, while maintaining a relatively low computational burden. Current research incorporates Triplet Attention into the YOLOv8n Backbone, enriching the model's information exchange with auxiliary keys and bolstering its ability to discern intricate structures[26].
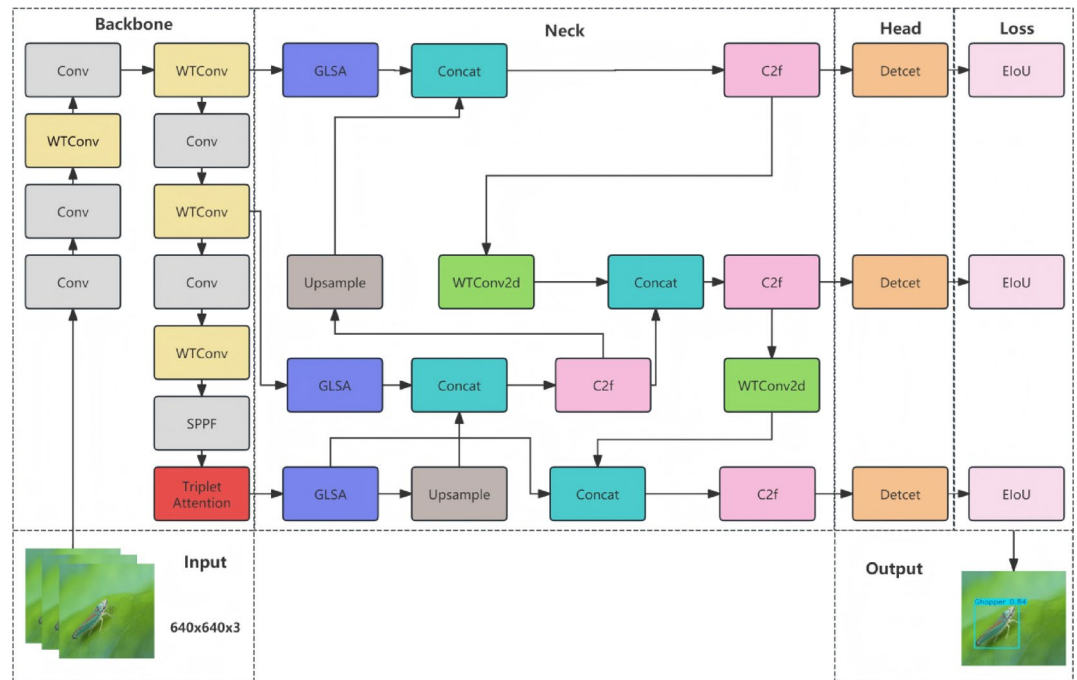
Fig. 3. Structure diagram of YOLO-DP. This figure illustrates the structure of YOLO-DP, including the Backbone, the Neck, the Head, the Loss, the Input and the Output. The backbone consists of Conv: A standard convolutional layer for extracting features, WTConv: Improved convolutional layers, GLSA: the Global to Local Spatial Aggregation) module, Concat: Feature stitching operation to fuse features at different levels, C2f: Feature Processing Module, Detect: The detection head is used to generate target detection results, and EloU:Loss function. The input and output image size is $640 \times 640 \times 3$.
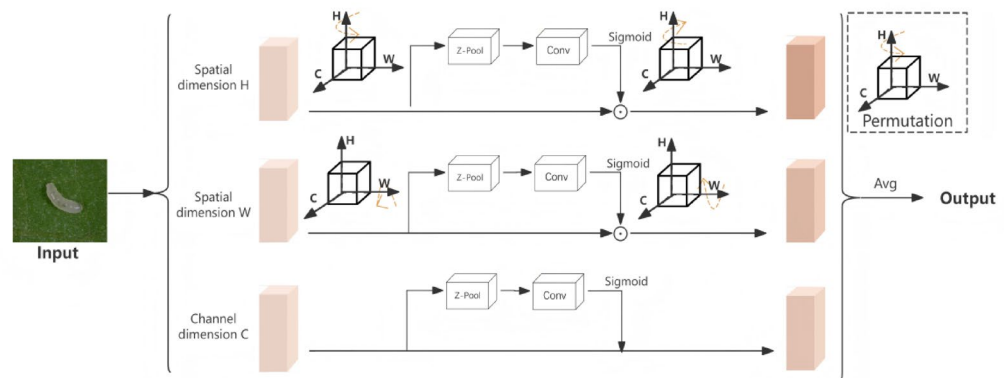

Fig. 4. Diagram of cross-dimensional interaction in triplelet attention. In this figure Triplet Attention consists of three parallel branches. Two of these branches are responsible for capturing cross-dimensional interactions between the channel (C) and spatial dimensions (H or W). The last branch is used to construct Spatial Attention. The outputs of the three branches are aggregated using averaging.

Triplet Attention's primary concept is Cross-dimensional Interaction: Triplet Attention mechanism adeptly captures the interplay between the channel (C) and spatial dimensions (height H and width W) via three distinct branches. Next, we have the following Rotation Operation: This operation adeptly captures dependencies across various dimensions without the need for dimensionality reduction, thereby reducing the parameter count while preserving the integrity of the information. Lastly, the Residual Connection plays a pivotal role: The output of each branch is integrated with the input through a Residual Connection, which bolsters the model's learning capacity. Figure 4 is an illustration of the core mechanism of Triplet Attention, Cross-dimensional Interaction.

Let the input feature map be χ, with dimensions $C \times H \times W$. The formulas applied range from formula 1 to 10. First Branch (Channel Attention): Compress channel information into two dimensions through $ZPool$ which includes max pooling and average pooling:

$$\chi_1^* = ZPool\left(\chi\right) \tag{1}$$

Generate attention weights through convolution and activation functions, where σ is the Sigmoid activation function, and ψ_1 is the convolution operation: Output, where $\odot$ denotes element-wise multiplication:

$$\omega_1 = \sigma\left(\psi_1\left(\chi_1^*\right)\right) \tag{2}$$

$$\overline{\chi_1} = \chi \odot \omega_1 \tag{3}$$

Second Branch (Channel and Height Interaction): Permute the feature map along the height dimension, where χ_{perm2} is the permutation of χ along the height dimension:

$$\chi_2 = \chi_{perm2} \tag{4}$$

Generate attention weights through $ZPool$ and convolution operations:

$$\omega_2 = \sigma\left(\psi_2\left(\chi_2^*\right)\right) \tag{5}$$

Output the average of χ_2:

$$\overline{\chi_2} = \chi \odot \omega_2 \tag{6}$$

Third Branch (Channel and Width Interaction): Permute the feature map along the width dimension, where χ_{perm3} is the permutation of χ along the width dimension: Generate attention weights through $ZPool$ and convolution operations:

$$\chi_3 = \chi_{perm3} \tag{7}$$

$$\omega_3 = \sigma\left(\psi_3\left(\chi_3^*\right)\right) \tag{8}$$

Output the average of χ_3:

$$\overline{\chi_3} = \chi \odot \omega_3 \tag{9}$$

Finally is the output part:Average the outputs of the three branches:

$$y = \tfrac{1}{3}\left(\overline{\chi_1} + \overline{\chi_2} + \overline{\chi_3}\right) \tag{10}$$

Global to local spatial aggregation (GLSA)

The efficacy of YOLOv8n is significantly influenced by its Neck architecture, which is tasked with merging feature maps across various scales to distill pertinent spatial details. Conventional Neck structures might fall short in effectively capturing both global and local features, particularly when encountering objects with substantial scale disparities. To address this challenge, we propose incorporating the Global to Local Spatial Aggregation (GLSA) module within the Neck of YOLOv8n, thus bolstering the model's capacity to discern global and local features, Fig. 5 is the overview of GLSA[27].

The GLSA module refines feature representation by isolating and combining global and local spatial features. This module is composed of two primary components: Global Spatial Attention (GSA) and Local Spatial Attention (LSA). The GSA component utilizes a self-attention mechanism to capture global dependencies, whereas the LSA component concentrates on the extraction of local features, which is crucial for the identification of small objects. The formulas applied range from formula 11 to 18.

Firstly, the Backbone of YOLOv8n was used to extract the multiscale feature map F_1', F_2', F_3'. Apply the GLSA module to each feature map to extract global and local features, where F_{gi} and F_{li} represent the global and local features of the ith feature map, respectively:

$$F_{gi} = GSA\left(F_i\right) \tag{11}$$

$$F_{li} = LSA\left(F_i\right) \tag{12}$$

Fusion of global and local features for richer representation of features:

$$F_i' = F_{gi} + F_{li} \tag{13}$$

The multi-head attention mechanism is applied to further fuse the features:

$$F_{multi} = MultiHead\left(F_1', F_2', F_3'\right) \tag{14}$$

Key formulas in the GLSA module is Global Spatial Attention (GSA):

$$AttG\left(F_i\right) = Softmax\left(Transpose\left(C_{1\times1}\left(F_i\right)\right)\right) \tag{15}$$

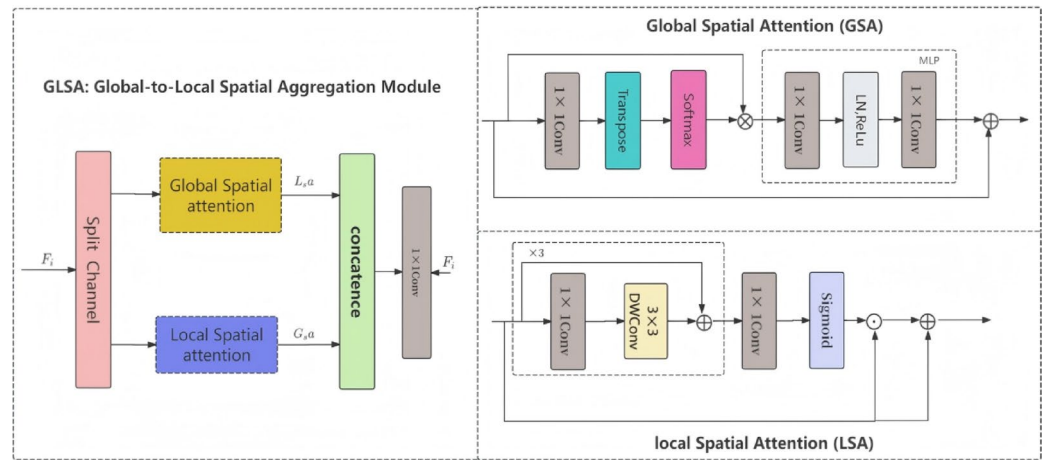$$GSA\left(F_i\right) = MLP\left(AttG\left(F_i\right) \otimes F_i\right) + F_i \tag{16}$$

Fig. 5. Overview of the global-to-local spatial aggregation module (GLSA). This figure shows the module consists of two parallel layers: a GSA (Global Self-Attention) layer, which focuses on the content of the pixels based solely on their content and a GLA (Global Location Attention) layer, which focuses on the spatial location of the pixels. The output of the module is the sum of the outputs from these two layers.

Local Spatial Attention (LSA):

$$AttL\left(F_i\right) = \sigma\left(C_{1\times1}\left(F_c\left(F_i\right)\right) + F_i\right) \tag{17}$$

$$LSA\left(F_i\right) = AttL\left(F_i\right) \odot F_i + F_i \tag{18}$$

Among them, $C_{1\times1}$ represents 1×1 convolution, $\otimes$ represents matrix multiplication, $\odot$ represents element-by-element multiplication, σ represents the sigmoid activation function, and MLP represents multilayer perception.

Wavelet transform convolution (WTConv)

The C2f-BottleNeck module in YOLOv8n requires feature fusion operation, which increases the computational complexity of the model, leads to increase of model training and inference time, and requires more storage space and computational resources. So we introduced Wavelet Transform convolution (WTConv) which is essentially a method of replacing traditional convolution kernel with wavelet basis function for signal or image processing. It replaces the fixed convolution kernel of traditional convolutional neural network with wavelet basis which has clear mathematical significance, so as to extract features in multi-scale and multi-direction, and combine the advantages of convolution's local receptive field and wavelet's multi-resolution analysis. Figure 6 is the WTConv operation example.

In this research, the WTConv module toke the place of the C2f-BottleNeck architecture of YOLOv8n, substituting the existing deep convolutional layer. This enhancement enables the current model to not only encompass a broader spectrum of contextual information but also to react more adeptly to low-frequency features, a crucial aspect for object detection tasks. Specifically, the WTConv module augments the model's sensitivity to shape via wavelet transformation, thereby enhancing the detection capabilities for small objects and those situated within intricate backgrounds. The introduction of WTConv modules can enhance network performance, especially has a low computing cost in terms of floating point operations[28].

The formula for WTConv can be represented by the following key steps:

Wavelet Transform (WT), where $X_{LL}^0 = X$ is the input to this layer, X_H^i rep-resents all three high-frequency plots of level i:

$$X_{LL}^i, X_H^i = WT\left(X_{LL}^{i-1}\right) \tag{19}$$

In the wavelet domain convolution, X is the input tensor, and W is a weighted tensor of a $k \times k$ deep convolution kernel with four times the number of input channels. This operation not only separates the convolution between the frequency components, but also allows the smaller convolution kernel to operate within a larger area of the original input, increasing the receptive field relative to the input:

$$Y = IWT\left(Conv\left(W, WT\left(X\right)\right)\right) \tag{20}$$

Cascading wavelet decomposition, where X_{LL}^i and X_H^i are wavelet transform results of the second order:

$$Y_{LL}^i, Y_H^i = Conv\left(W^i, \left(X_{LL}^i, X_H^i\right)\right) \tag{21}$$
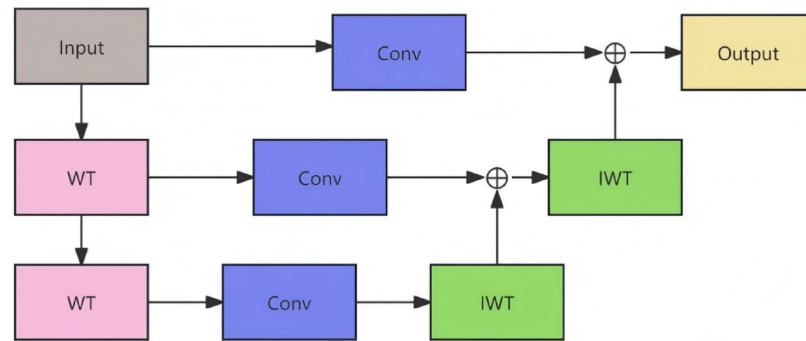
Fig. 6. WTConv operation example. This figure shows the WTConv operation separates the convolution across frequency components and also allows smaller convolution kernels to operate over a larger region of the original input, thereby increasing the receptive field relative to the input. WT (Wavelet Transform) aims to expand the receptive field of convolution through wavelet transformation, and IWT (Inverse Wavelet Transform) is the linear inverse of WT, which can reconstruct the convolution results back to the original space without loss.

Combination of different frequency outputs, this results in the summation of different levels of convolution, where Z^i is the aggregate output from level i and beyond:

$$Z^i = IWT\left(Y_{LL}^i + Z^{i+1}, Y_H^i\right) \tag{22}$$

Enhanced intersection over union (EIoU)
The YOLOv8n model adopts the CIoU (complete intersection over union) Loss function, which is composed of the original IoU (intersection over union), the aspect ratio calculation, and the distance from the center point, which can evaluate the similarity between the prediction box and the actual box for bounding box regression, but when the regression bounding box is performed, if the height-width ratio of the predicted box and the actual box shows a linear correlation, The penalty term of the CIoU loss function may be-come invalid and become zero[29].

The EIoU loss function in this paper is further improved on the basis of the CIoU loss function, which not only considers the distance between the center point of the prediction box and the real box, but also considers the difference between the aspect ratio of the prediction box and the aspect ratio of the real box. This improvement allows EIoU to decouple geometric factors and optimize more accurately and evaluate the accuracy of the prediction frame more comprehensively, especially when the prediction frame has less overlap with the real box, EIoU can provide more efficient gradient information and accelerate the convergence of the model. Figure 7 is EIoU loss function schematic diagram.

The EIoU loss function is calculated as follow:

$$EIoU = 1 - \left(\frac{intersection}{union}\right) + \frac{\left(w^{pred} - w^{gt}\right)^2}{C_w^2} + \frac{\left(h^{pred} - h^{gt}\right)^2}{C_h^2} + \frac{\rho^2\left(b^{pred}, b^{gt}\right)}{c^2} \tag{23}$$

In this formula, $intersection$ is the area where the prediction box intersects with the real box, $union$ is the union area of the predicted box and the true box. $\rho\left(b^{pred}, b^{gt}\right)$ is euclidean distance between box centers. c is the diagonal length of the smallest enclosing box. $w^{pred} - w^{gt}$ is the square of the difference between the width of the predicted box and the width of the true box. $h^{pred} - h^{gt}$ is the square of the difference between the height of the predicted box and the height of the true box. c_w is the width of the minimum additive box that contains the predicted box and the real box. c_h is the height of the smallest encore box that contains the prediction box and the real box.

Model and and result analysis
Evaluation indicators
In order to test the performance of the improved network model, frames per second at the time of verification, the size of the model, Precision (P), Recall (R) , the average precision of all detection categories when at 50% overlap between the prediction box and the real box (mAP50) and the average precision (mAP95) of all detection categories with a stride length of 0.05 when the overlap between predicted and real boxes is between 50 and 95% are used as the final evaluation indexes of the model performance. The calculation formula can be expressed as:

$$P = \frac{TP}{TP+FP} \tag{24}$$
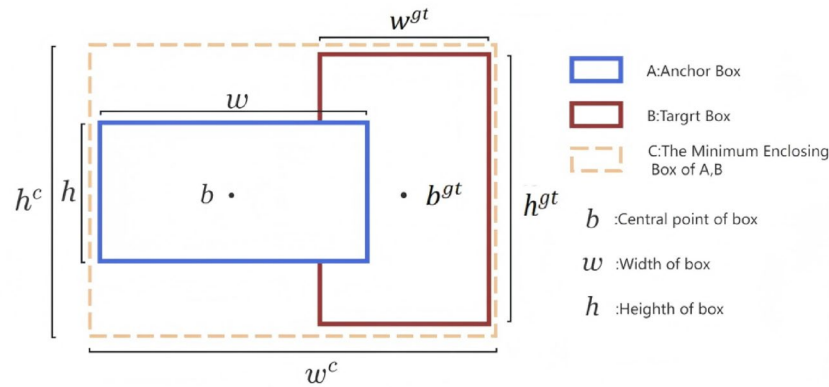
$$R = \frac{TP}{TP+FN} \tag{25}$$

Fig. 7. EIoU loss function schematic diagram. This figure shows EIoU Loss focuses on higher-quality anchor boxes by decomposing the aspect ratio influence factor of the prediction box and the ground-truth box, and separately calculating the differences in width and height.

Model	FPS	Size/MB	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)
TOOD	50.6	246.0	71.0	61.3	66.3	47.6
Faster R-CNN	52.8	1230.0	65.4	60.5	63.6	44.1
RT-DETR	8.1	38.6	80.6	74.0	74.3	46.4
YOLOv10n	166.9	5.5	71.0	50.4	58.7	38.1
YOLOv8n	268.7	6.0	79.0	70.7	74.9	46.1
YOLO-DP	138.1	11.8	**80.9**	**74.4**	**77.8**	**50.1**

Table 3. Experimental results of the comparative test. Significant values are in [bold].

$$AP = \int_0^1 P(R)\, dR \tag{26}$$

$$mAP50 = \frac{\sum_{j=1}^k AP_i}{k(classes)}\,(IoU = 0.5) \tag{27}$$

$$mAP95 = \frac{\sum_{j=1}^k \sum_{t=0}^9 AP_i}{10k(classes)}\,(IoU = 0.5 + 0.05t) \tag{28}$$

Results analysis
Comparative experiment results and analysis
To verify the effect of the improved model, selected TOOD(Task-aligned One-stage Object Detection)[30], typical representative of the two-stage object detection model Faster R-CNN[31], RT-DETR (Real-Time Detection Transformer)[32], YOLOv8n, YOLOv10n[33] and YOLO-DP for comparative test, the comparison results in Table 3. It can be seen that YOLO-DP is better than other popular models in the detection of rice diseases and pests, improves by 2.9% based on the original YOLOv8n mAP50 and 4% of mAP95.

Figure 8 illustrates the results of detecting rice pests and diseases across six different models in various scenarios and times, including multiple detection targets, close-range, long-distance, direct sunlight, and nighttime. In conditions of long-distance or direct sunlight, RT-DETR and YOLOv10n may fail to detect rice pests and diseases. TOOD and Faster R-CNN have relatively lower accuracy, whereas YOLO-DP, based on YOLOv8n, offers higher precision in identifying rice pests and diseases, making it more effective for detection.

Figure 9 is a curve plot of the training epochs of six different models compared to the values of mAP50. As can be seen from the curve chart, the mAP50 value of YOLO-DP is higher than that of other models, showing the superiority of the model. Figure 10 shows the comparison of Precision-Recall Curve between YOLOv8n and YOLO-DP, which shows that the model has a good target performance and improve for fifteen types of rice diseases and pests.

Figure 11 is a visual system which loaded YOLO-DP model to capture images of rice diseases and pests. At the same time, the system supports video detection and camera detection, and displays the location, quantity, reliability, and the consumed time of detecting targets.

Ablation experiment
In order to verify the effectiveness of the improved algorithm, four networks were designed for ablation experiments, all using the same rice diseases and pests data set, batches, and training period. The experimental
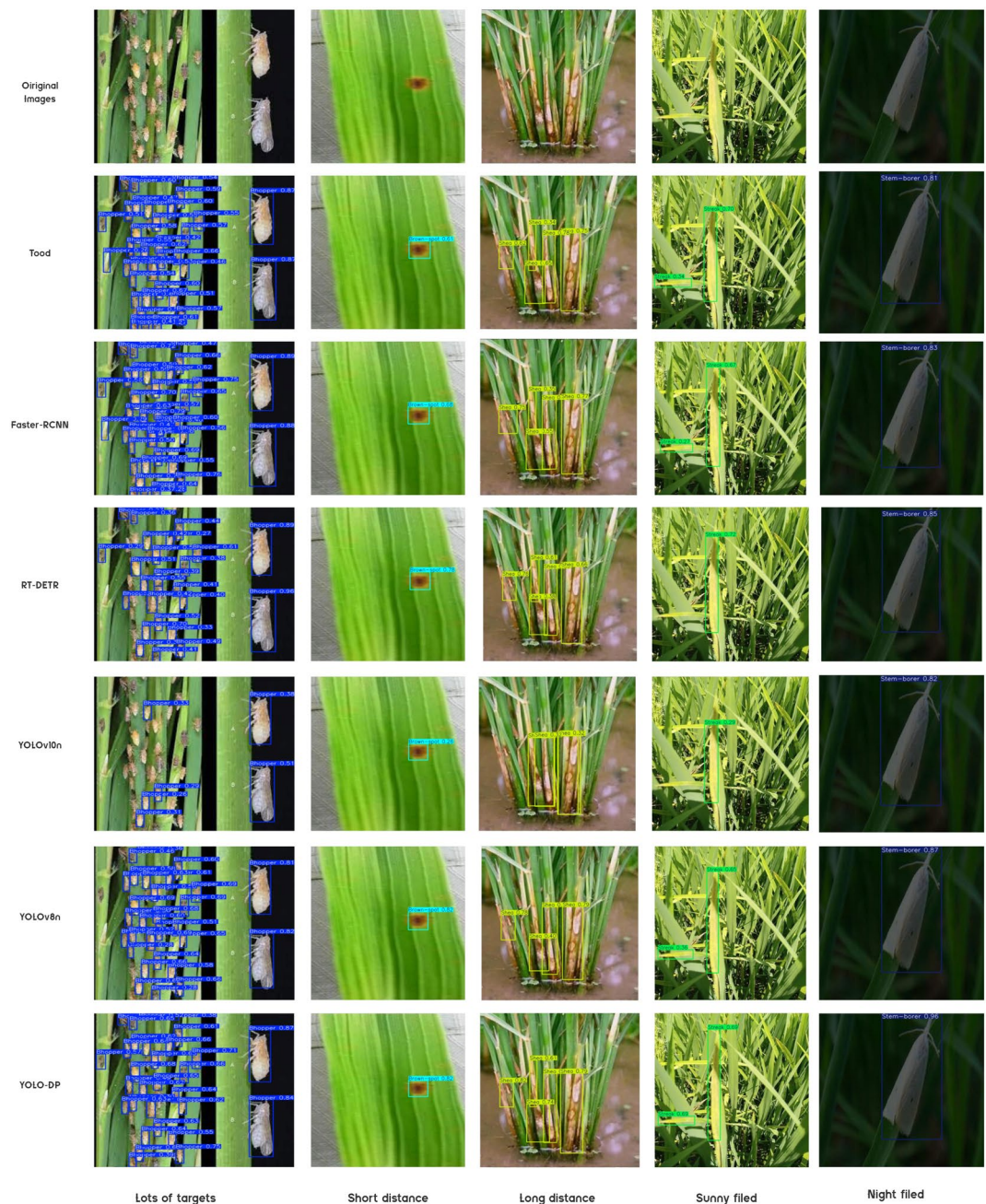
Fig. 8. Comparison of six models for rice diseases and pests detection. This figure is the detection comparison chart of the six models shows the average precision of all detection categories when at least 50% overlap between the prediction box and the real box of rice diseases and pests detection under various conditions. It can be seen that YOLO-DP performs better.

results are shown in Table 4, except the Average Accuracy of YOLOv8n + EIoU, value decreased by 0.66%, each other improvement point was improved.

Comparison of different attention mechanisms

To comprehensively evaluate the performance of Triplet Attention in YOLOv8n model, this study conducted comparative experiments including three types of attention mechanism: YOLOv8n + AgentAttention[34], YOLOv8n + LocalWindowAttention[35], and YOLOv8n + Triplet Attention. The three attention mechanisms were added separately to the tail of the network Backbone, the experimental results are shown in Table 5.

According to the experiment, Triplet Attention is superior to other attention mechanisms when used to detect diseases and pests in rice.
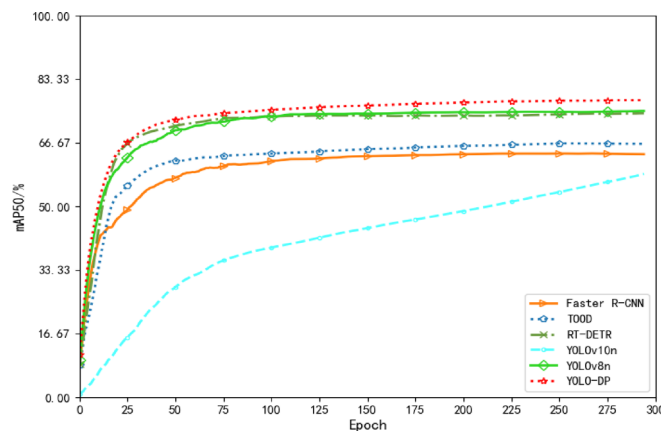
Fig. 9. Curve plot of the training epochs against the values of mAP50. This figure is a curve plot of the training epochs of six different models compared to the values of mAP50. It can be seen that YOLO-DP performs better.
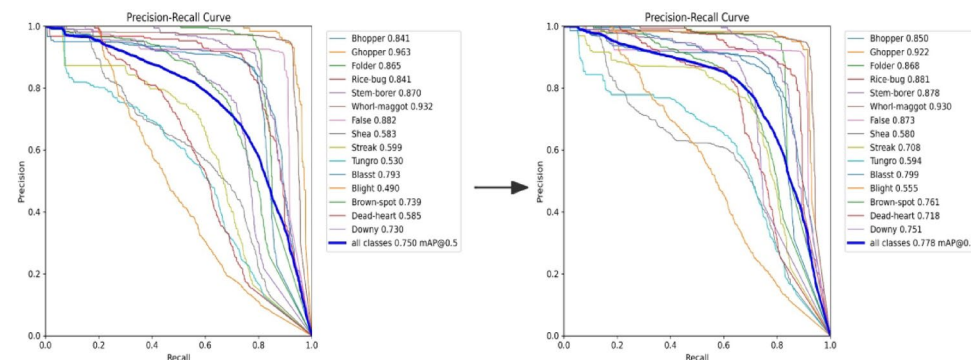

Fig. 10. Precision-recall curve comparison between YOLOv8n and YOLO-DP. This figure shows the values of mAP50 for 15 kinds of rice diseases and insect pests before and after the model improvement, which proved that the model improved the detection of 15 kinds of rice diseases and pests.

Comparison of different loss functions

To comprehensively evaluate the performance of EIoU in enhancing the YOLOv8n model, this study conducted comparative experiments involving four loss functions: YOLOv8n + EMASlideLoss[36], YOLOv8n + SlideLoss[37], YOLOv8n + DIoU[38], and YOLOv8n + EIoU. The objective was to analyze the performance differences among these various loss functions in the context of rice diseases and pests detection tasks, the experimental results are shown in Table 6.

Because the indicators of YOLOv8n + EIoU are higher than other loss functions, according to the experimental results, YOLOv8n + EIoU is more suitable to detect rice diseases and pests than other loss functions.

Heat map analysis

Using Grad-CAM (Gradient-weighted Class Activation Mapping)[39] visual heat maps to visualize the features of different rice diseases and pests output by the YOLO-DP model, the results are shown in Fig. 12. This figure can visually display the level of regional attention, with more obvious features corresponding to higher heat at the respective positions, the confidence and prediction boxes for rice diseases and pests detection are also shown.

As can be seen in Fig. 12, compared to YOLOv8n, YOLO-DP can locate the characteristics of diseases and pests and perform accurate identification more accurately, the confidence level of detection is also higher and also reduces the impact of complex backgrounds on detection.

Analysis of other data set

To verify the generalizability of the method proposed in this study, comparative experiments were conducted on other datasets using YOLOv8n and YOLO-DP. The experiment utilized the 2020 Rice Leaf Disease Image Samples[40]. The data set contains 5932 number images includes four kinds of Rice leaf diseases i.e. Bacterial blight, Blast, Brown Spot and Tungro. The experiment selected 3849 images and divided them into training, validation, and test sets in a 7:2:1 ratio, with experimental environment and parameter settings consistent with those described in "Results analysis" section. The results are shown in Table 7. Compared to YOLOv8n, YOLO-DP improved the average accuracy and recall rates by 1.3% and 0.7%. There was a significant 1.4% improvement
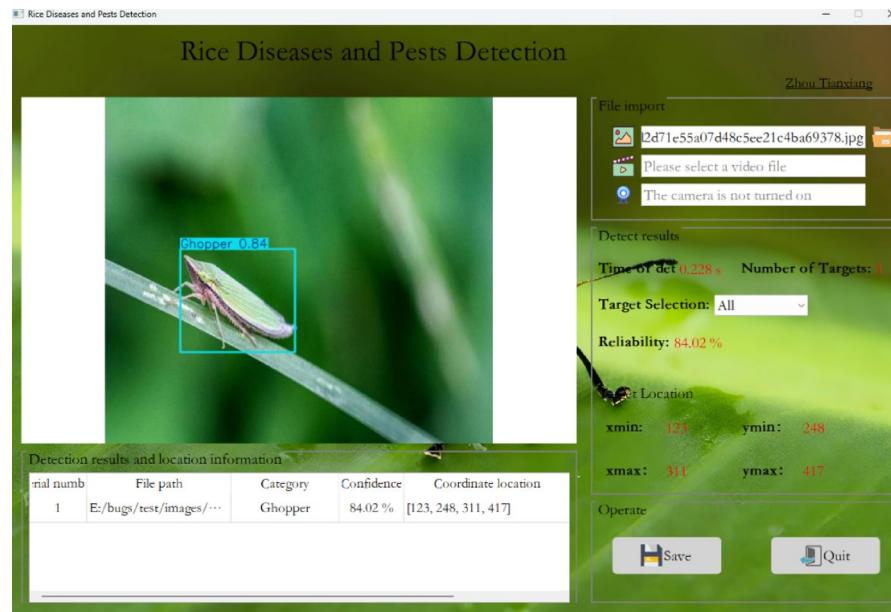
Fig. 11. Visual system of rice diseases and pests detection. This figure is a visual system which loaded YOLO-DP model to capture images of rice diseases and pests. At the same time, the system supports video detection and camera detection, and displays the location, quantity, reliability, and the consumed time of detecting targets. It can be seen that YOLO-DP performs well.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)
YOLOv8n	79.0	70.7	74.9	46.1
YOLOv8n + EIoU	78.3	**72.5**	**76.2**	**47.7**
YOLOv8n + Triplet Attention	**79.0**	**72.3**	**76.4**	**47.4**
YOLOv8n + WTConv	**79.2**	**71.4**	**75.7**	**48.0**
YOLOv8n + GLSA	**80.6**	**71.8**	**76.2**	**48.9**

Table 4. Experimental results of the ablation experiment. Significant values are in [bold].

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)
YOLOv8n	79.0	70.7	74.9	46.1
YOLOv8n + AgentAttention	77.8	72.6	76.0	47.0
YOLOv8n + LocalWindowAttention	77.8	71.7	75.6	46.6
YOLOv8n + Triplet Attention	**79.0**	**72.3**	**76.4**	**47.4**

Table 5. Experimental results of different attention mechanisms. Significant values are in [bold].

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)
YOLOv8n	79.0	70.7	74.9	46.1
YOLOv8n + EMASlideLoss	79.5	71.2	75.5	46.8
YOLOv8n + SlideLoss	78.4	71.9	75.9	47.2
YOLOv8n + DIoU	78.6	71.0	75.0	47.2
YOLOv8n + EIoU	78.3	**72.5**	**76.2**	**47.7**

Table 6. Experimental results of different loss functions. Significant values are in [bold].

Fig. 12. Heat map analysis between YOLOv8n and YOLO-DP. This figure shows the Grad-CAM technique was used to analyze and compare the performance of the original YOLOv8n model and YOLO-DP in object detection tasks. Heatmap areas from red to yellow represent the primary focus of the model, while blue areas indicate lower attention. YOLO-DP's attention is more focused on the center of the diseases and pests, demonstrating more precise feature extraction.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP95 (%)
YOLOv8n	89.2	87.4	91.1	72.0
YOLO-DP	**90.5**	**88.1**	**92.5**	**72.1**

Table 7. Experimental results of 2020 rice leaf disease image samples. Significant values are in [bold].

in the mean Average Precision over the original YOLOv8n model, indicating higher detection accuracy of YOLO-DP.

Conclusions

Given the current lack of effective methods for simultaneously detecting rice diseases and pests, this study innovatively constructed a dataset that meticulously includes images of the fifteen most common rice diseases and pests. The YOLO-DP model was then introduced, which significantly improved the accuracy of detection under various lighting conditions, distances, and target quantities. By comparing with multiple mainstream object detection models, attention mechanisms, and loss functions, and through ablation studies, the effectiveness of the rice disease and pest detection model was validated. Finally, a rice disease and pest detection system was developed, supporting image, video, and camera detection.

Statistical analysis of the experimental results shows that the YOLO-DP model exhibits stable and excellent performance in rice disease and pest detection tasks. In comparative experiments with the original YOLOv8n model, it achieves a 1.9% increase in precision, a 3.7% increase in recall, a 2.9% increase in mAP50, and a 4.0% increase in mAP95, reaching 90.5, 88.1, 92.5, and 72.1 respectively. This indicates that the improvement of the YOLO-DP model is not due to accidental fluctuations but a stable performance enhancement achieved through structural optimization. In addition, the model's decision-making process is analyzed using the Grad-CAM visualization tool, which further verifies the accuracy and consistency of the model in capturing disease and pest features, providing quantitative support for the interpretability of the results.

In conclusion, YOLO-DP model shows excellent performance in rice disease and pest detection tasks, providing strong technical support for intelligent agriculture and real-time disease monitoring, and providing the possibility to apply these advanced model technologies to other crops and diseases[41].

Future research could explore extending the YOLO-DP model to detect diseases in other crops, such as wheat and corn, to validate and optimize its applicability and effectiveness in a broader range of crop diseases, more specialized models for small-scale diseases and pests detection will be introduced into the experimental

comparison. For extreme scenarios that are not covered in the field, such as blurred images and dew reflection on foggy days, the dataset is planned to be expanded, and adversarial training strategies are introduced to improve the model's resistance to unknown interference[42]. Additionally, incorporating multi-modal image detection technology, which utilizes information from various imaging modes (such as visible light, infrared, and hyperspectral images), can enrich the model's input dimensions, enhance the distinguishability of pest and disease features, and provide forward-looking ideas for developing more precise and versatile intelligent monitoring systems for crop pests and diseases[43]. Finally, the model is connected to agricultural Internet of Things (IoT) devices, through model quantification and knowledge distillation, it is adapted to low-computing power devices and the lightweight YOLO-DP model is deployed on drones and field edge computing terminals, enabling full-process automation of "real-time image collection—local rapid detection"[44].

Data availability

The data supporting this study's findings are available from the corresponding author upon reasonable request. The data set and code for this paper have been open-sourced and are available at:https://github.com/ztx1448368118/yolodp

References

1. Li, R. et al. Predicting rice diseases using advanced technologies at different scales: Present status and future perspectives. *Abiotech* **4**, 359–371 (2023).
2. Tussupov, J. et al. Analysis of formal concepts for verification of pests and diseases of crops using machine learning methods. *IEEE Access* **12**, 19902 (2024).
3. Upadhyay, N. & Gupta, N. Detecting fungi-affected multi-crop disease on heterogeneous region dataset using modified ResNeXt approach. *Environ. Monit. Assess.* **196**, 610 (2024).
4. Upadhyay, N. & Bhargava, A. Artificial intelligence in agriculture: Applications, approaches, and adversities across pre-harvesting, harvesting, and post-harvesting phases. *Iran J. Comput. Sci.* **8**, 1–24 (2025).
5. Zheng, Q. et al. Remote sensing monitoring of rice diseases and pests from different data sources: A review. *Agronomy* **13**, 1851 (2023).
6. Chen, X. et al. Design and test of target application system between rice plants based on light and tactile sensing. *Crop Prot.* **182**, 106722 (2024).
7. Zhang, J. et al. Monitoring plant diseases and pests through remote sensing technology: A review. *Comput. Electron. Agric.* **165**, 104943 (2019).
8. Rahman, C. et al. Identification and recognition of rice diseases and pests using convolutional neural networks. *Biosys. Eng.* **194**, 112–120 (2020).
9. Li, D. et al. A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors* **20**, 578 (2020).
10. Mukherjee, R. et al. Rice leaf disease identification and classification using machine learning techniques: A comprehensive review. *Eng. Appl. Artif. Intell.* **139**, 109639 (2025).
11. Prajapati, H., Shah, J. & Dabhi, V. Detection and classification of rice plant disease. *Intell. Decis. Technol.* **11**, 357–373 (2017).
12. Zhang, Y. et al. Resvit-rice: A deep learning model combining residual module and transformer encoder for accurate detection of rice diseases. *Agriculture* **13**, 1264 (2023).
13. Uddin, M. et al. E2ETCA: End-to-end training of CNN and attention ensembles for rice disease diagnosis. *J. Integr. Agric.* (2024).
14. Rajasekhar, V., Arulselvi, G. & Babu, K. Design an optimization based ensemble machine learning framework for detecting rice leaf diseases. *Multimed. Tools Appl.* **83**, 1–24 (2024).
15. Sangaiah, A. et al. UAV T-YOLO-rice: An enhanced tiny Yolo networks for rice leaves diseases detection in paddy agronomy. *IEEE Trans. Netw. Sci. Eng.* **11**, 5201 (2024).
16. He, K., Zhang, X., Ren, S. & Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1904–1916 (2015).
17. Gao, W. et al. Intelligent identification of rice leaf disease based on YOLO V5-EFFICIENT. *Crop Prot.* **183**, 106758 (2024).
18. Diwan, T., Anirudh, G. & Tembhurne, J. Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Sensors* **24**, 893 (2024).
19. Song, H., Yan, Y., Deng, S., Jian, C. & Xiong, J. Innovative lightweight deep learning architecture for enhanced rice pest identification. *Phys. Scr.* **99**, 096007 (2024).
20. Zhang, X. et al. LDConv: Linear deformable convolution for improving convolutional neural networks. *Image Vis. Comput.* **149**, 105190 (2024).
21. Wu, T. et al. Cgnet: A light-weight context guided network for semantic segmentation. *IEEE Trans. Image Process.* **30**, 1169–1179 (2020).
22. Trinh, D. et al. Alpha-EIOU-YOLOv8: An improved algorithm for rice leaf disease detection. *AgriEngineering* **6**, 302–317 (2024).
23. He, J. et al. Alpha-IoU: A family of power intersection over union losses for bounding box regression. *Adv. Neural. Inf. Process. Syst.* **34**, 20230–20242 (2021).
24. Wang, J., Jin, L., Li, Y. & Cao, P. Application of end-to-end perception framework based on boosted DETR in UAV inspection of overhead transmission lines. *Drones* **8**, 545 (2024).
25. Upadhyay, N. & Gupta, N. SegLearner: A segmentation based approach for predicting disease severity in infected leaves. *Multimed. Tools Appl.* 1–24 (2025).
26. Guo, M. et al. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media* **8**, 331–368 (2022).
27. Hossain, M. et al. DeepPoly: Deep learning-based polyps segmentation and classification for autonomous colonoscopy examination. *IEEE Access* **11**, 95889–95902 (2023).
28. Guo, Z. et al. DSNet: A novel way to use atrous convolutions in semantic segmentation. *IEEE Trans. Circuits Syst. Video Technol.* **35**, 3679 (2024).
29. Rahman, M. M., Tan, Y., Xue, Ji. & Lu, K. Notice of violation of IEEE publication principles: Recent advances in 3D object detection in the era of deep neural networks: A survey. *IEEE Trans. Image Process.* **29**, 2947–2962 (2019).
30. Terven, J., Crdova-Esparza, D. & Romero-Gonzlez, J. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach. Learn. Knowl. Extr.* **5**, 1680–1716 (2023).
31. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2016).

32. Li, H. et al. Delving into the devils of bird's-eye-view perception: A review, evaluation and recipe. *IEEE Trans. Pattern Anal. Mach. Intell.* **46**, 2151 (2023).

33. Wang, A. et al. Yolov10: Real-time end-to-end object detection. Preprint at arXiv:2405.14458 (2024).

34. Guan, Y. et al. CTWheatNet: Accurate detection model of wheat ears in field. *Comput. Electron. Agric.* **225**, 109272 (2024).

35. Jamali, A., Roy, S. K., Bhattacharya, A. & Ghamisi, P. Local window attention transformer for polarimetric SAR image classification. *IEEE Geosci. Remote Sens. Lett.* **20**, 15 (2023).

36. Yu, Z. et al. Yolo-facev2: A scale and occlusion aware face detector. *Pattern Recogn.* **155**, 110714 (2024).

37. Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **11**, 677 (2023).

38. Lu, S., Lu, H., Dong, J. & Wu, S. Object detection for UAV aerial scenarios based on vectorized IOU. *Sensors* **23**, 3061 (2023).

39. Selvaraju, R. R. et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vision* **128**, 336–359 (2020).

40. Baidu. (n.d.). Rice disease pictures dataset-Paddle AI Studio Galaxy Community. Retrieved from https://aistudio.baidu.com/aistudio/datasetdetail/266101

41. Upadhyay, N. & Gupta, N. Diagnosis of fungi affected apple crop disease using improved ResNeXt deep learning model. *Multimed. Tools Appl.* **83**, 64879–64898 (2024).

42. Selvaraju, R. R. et al. Mango crop maturity estimation using meta-learning approach. *J. Food Process Eng* **47**, e14649 (2024).

43. Li, R. et al. Identification of cotton pest and disease based on CFNet-VoV-GCSP-LSKNet-YOLOv8s: a new era of precision agriculture. *Front. Plant Sci.* **15**, 1348402 (2024).

44. Qin, W. et al. Automated lepidopteran pest developmental stages classification via transfer learning framework. *Environ. Entomol.* **53**, 1062–1077 (2024).

Author contributions

T.Z: Performed the majority of the experiments, contributed to the conception and design of the study, wrote the initial draft of the manuscript, and assisted in revisions, played a key role in analyzing data, and significantly contributed to writing and revising the manuscript. L.W: Contributed to the conception and design of the study, supervised and guided the experiments. T.Z: Assisted with data analysis and interpretation, contributed to manuscript revisions, and approved the final version. L.W: Contributed to the development of the methodology, assisted with the experiments, and reviewed and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.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.