

Article

A Study on Maize Leaf Pest and Disease Detection Model Based on Attention and Multi-Scale Features

Jie Kang *, Wanhu Zhang *, Yu Xia and Wenbo Liu

School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China; yuxia@sust.edu.cn (Y.X.); wblu@sust.edu.cn (W.L.)

* Correspondence: kangjie@sust.edu.cn (J.K.); 210611015@sust.edu.cn (W.Z.)

Abstract: The detection and accurate positioning of agricultural pests and diseases can significantly improve the effectiveness of disease and pest control and reduce the cost of prevention and control, which has become an urgent need for crop production. Aiming at the low precision of maize leaf pest and disease detection, a new model of maize leaf pest and disease detection using attention mechanism and multi-scale features was proposed. Our model combines a convolutional block attention module (CBAM) with the ResNet50 backbone network to suppress complex background interference and enhance feature expression in specific regions of the maize leaf images. We also design a multi-scale feature fusion module that aggregates local and global information at different scales, improving the detection performance for objects of varying sizes. This module reduces the number of parameters and enhances efficiency by using a lightweight module and replacing the deconvolutional layer. Experimental results on a natural environment dataset demonstrate that our proposed model achieves an average detection accuracy of 85.13%, which is 9.59% higher than the original CenterNet model. The model has 24.296 M parameters and a detection speed of 23.69 f/s. Compared with other popular models such as SSD-VGG, YOLOv5, Faster-RCNN, and Efficientdet-D0, our proposed model demonstrates superior performance in the fast and accurate detection of maize leaf pests and diseases. This model has practical applications in the identification and treatment of maize pests and diseases in the field, and it can provide technical support for precision pesticide application. The trained model can be deployed to a web client for user convenience.

Keywords: maize leaf; disease and pest detection; CenterNet; multi-scale feature fusion; convolution attention module



Citation: Kang, J.; Zhang, W.; Xia, Y.; Liu, W. A Study on Maize Leaf Pest and Disease Detection Model Based on Attention and Multi-Scale Features. *Appl. Sci.* **2023**, *13*, 10441. <https://doi.org/10.3390/app131810441>

Academic Editor: Norman Q. Arancon

Received: 18 July 2023

Revised: 8 September 2023

Accepted: 14 September 2023

Published: 18 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Maize, one of China's most important food crops and industrial raw materials, was cultivated on 43,070 thousand hectares in 2022, which was 1.37 times larger than rice and 1.72 times larger than wheat. However, the growth of maize is often hindered by pests and diseases, posing a significant challenge for farmers [1]. The misuse of pesticides not only affects crop quality and yield but also leads to excessive residue in agricultural products and environmental pollution, posing a threat to the sustainable development of agriculture [2]. Common maize leaf diseases include northern leaf blight (NLB), northern leaf spot (NLS), and grey leaf spot (GLS) [3], as well as pests such as maize borers, aphids, red spiders, armyworms, and peach borers. These diseases and pests have a severe impact on maize production [4].

With the advancement of artificial intelligence, there has been a growing interest in utilizing deep learning techniques for crop pest detection [5]. Xie et al. [6] proposed a grape leaf disease detection model based on a modified Faster R-CNN, achieving an average accuracy of 81.1% and a detection speed of 15.01 f/s. Liu et al. [7] improved the YOLOv3 algorithm to detect tomato pests and diseases with a detection accuracy of 92.39 percent. Richey et al. [8] enhanced the YOLOv4 algorithm to detect northern leaf

blight in maize, achieving an average detection accuracy of 93.55% for this particular disease. Sun [9] proposed the MEAN-SSD algorithm based on the improved SSD algorithm for detecting apple leaf disease, with an average detection accuracy of 83.12%. Yang et al. [10] developed a maize stamen detection model by improving the CenterNet algorithm to realize maize growth monitoring and yield estimation, achieving a recognition accuracy of 92.4%. However, the CenterNet model had some instances of missing dense targets.

While Faster R-CNN and other two-stage methods exhibit high detection accuracy, they are computationally intensive and may not meet real-time requirements. On the other hand, one-stage algorithms like YOLO directly take the image as input and learn pixel-level features, but they often suffer from a high false detection rate, misclassifying some background areas as targets. Additionally, most of the leaf image datasets used in research are captured in laboratory environments with relatively simple backgrounds, which differ significantly from the complex backgrounds found in real environments, including various soils, vegetation, stones, and weeds. These varying background factors can interfere with plant disease detection and affect algorithm performance.

To address these issues, this paper proposes an improved maize leaf disease and pest detection model based on the CenterNet target detection algorithm. The model focuses on several common maize leaf diseases and pests in natural environments, including three diseases and five pests. Experimental comparisons with other detection algorithms are conducted to verify the effectiveness of the proposed method. Furthermore, the model is deployed on a web platform to provide farmers with a more convenient user experience.

2. Materials and Methods

2.1. Dataset

In this study, a total of 2775 images of maize leaf pests and disease were collected, including 724 disease images and 2051 insect pest images. The disease images included three categories of diseases: northern leaf blight (NLB), northern leaf spot (NLS), and grey leaf spot (GLS). The insect pest images included five types of pests: maize borer, aphid, red spider, armyworm, and peach borer. Figure 1 provides example images for various pests and diseases.

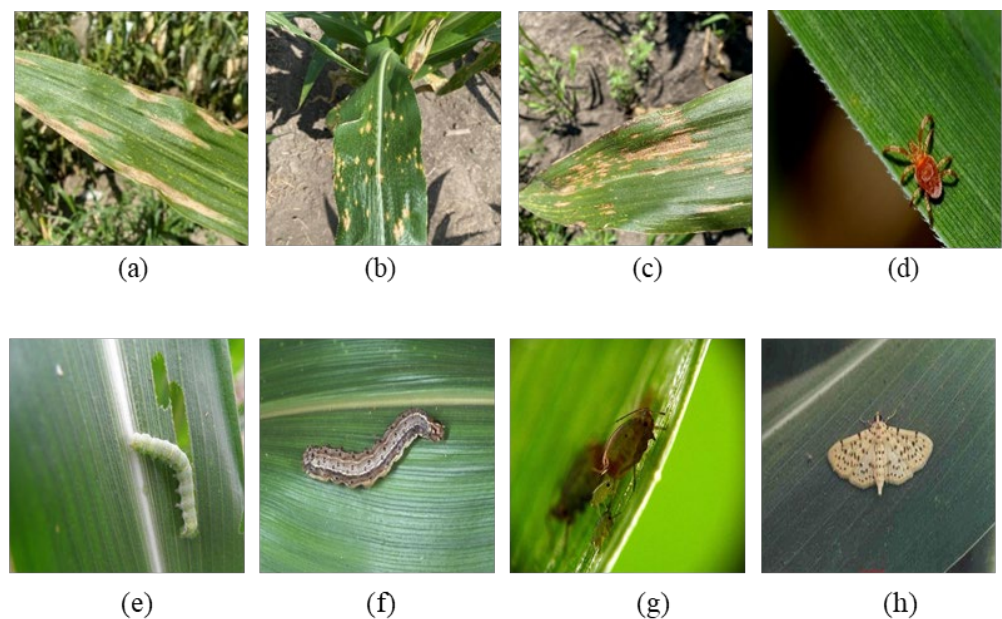


Figure 1. Images of some samples of pests and diseases in the dataset. (a) Northern leaf blight, (b) northern leaf spot, (c) gray leaf spot, (d) red spider, (e) maize borer, (f) armyworm, (g) aphids, (h) peach borer.

The dataset was annotated using the LabelImg image annotation tool, which generates an XML file containing object information that is saved in PASCAL VOC format along with the corresponding image file. The dataset was divided into an 80% training set, a 10% verification set, and a 10% test set. The original image was cropped and scaled by affine transformation to 512×512 pixels as input to the model. The distribution of the data samples is presented in Table 1.

Table 1. Data samples with the same statistics.

Pest and Disease Name	Sample Size
Northern leaf blight	249
Northern leaf spot	213
Gray leaf spot	262
Maize borer	424
Aphids	879
Red spider	160
Armyworm	206
Peach borer	382
Total sample size	2775

CenterNet model is an anchor-free box detection algorithm proposed by ZHOU et al. [11], which takes the object as a key point to estimate and then returns to other object attributes. First, the CenterNet model generates a feature map of the input RGB image through a feature extraction network and then sends the feature map to a detection network. The detection network consists of three branches that predict the heatmap of the object, the offset of the center point, and the width and height of the object.

2.2. Proposed Improve Centernet Model

CenterNet network adopts the “encoder–decoder” structure. After 32 times down-sampling of the model, the feature layer of high-level semantic information is obtained via deconvolution for object detection. However, this model does not consider the detection ability of multi-scale objects, and the detection ability of small objects is weak. In the process of feature extraction, multiple downsampling will cause the feature aggregation of small objects, leading to the problem of missing detection and wrong detection. Additionally, the different object types of maize leaf pests and diseases, as well as the variation in object size, create a significant increase in detection difficulty, given the small and dense characteristics of these objects.

To address these issues, we propose a maize leaf pest and disease detection model based on the CenterNet algorithm. As shown in Figure 2, the model consists of three main parts: backbone network, neck network and detection network. The backbone network uses ResNet50 to extract features and embeds the convolutional attention module (CBAM) in the second to fourth layers, which helps to learn the distribution law of features and improve the weights of semantic and location features for pests and diseases. The neck network uses multi-scale feature fusion module (MFF) to replace the deconvolutional layer of the original model. The module fuses the input feature maps using a weighted bidirectional feature fusion network and balances the contributions of feature maps with different resolutions in the output by adding additional weights. Finally, the feature maps at different scales are concatenated and fused with each other and fed into the detection network.

By improving the CenterNet model, the maize leaf pest detection model is able to better detect small objects, as well as improve detection efficiency and accuracy.

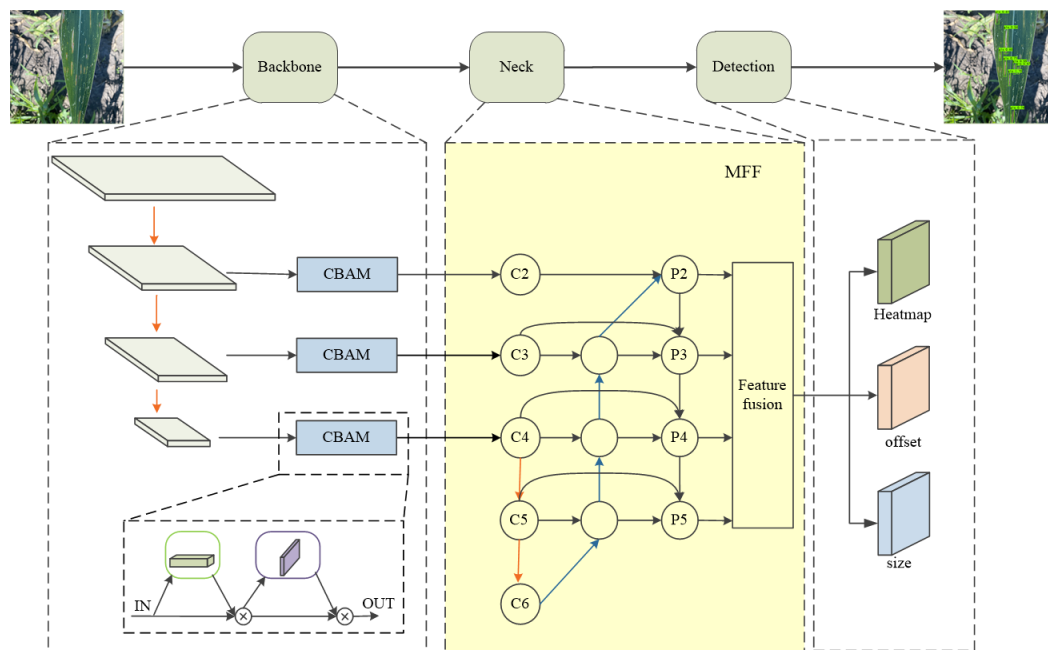


Figure 2. Improved CenterNet model for maize leaf pest and disease detection.

2.2.1. Convolutional Block Attention Module

To enhance feature expression in specific regions of the image, the attention module CBAM [12] is embedded in the backbone network. This module is a lightweight universal module that can be easily embedded into the network model and has little effect on the parameter scale of the model. The structure of the CBAM module is illustrated in Figure 3.

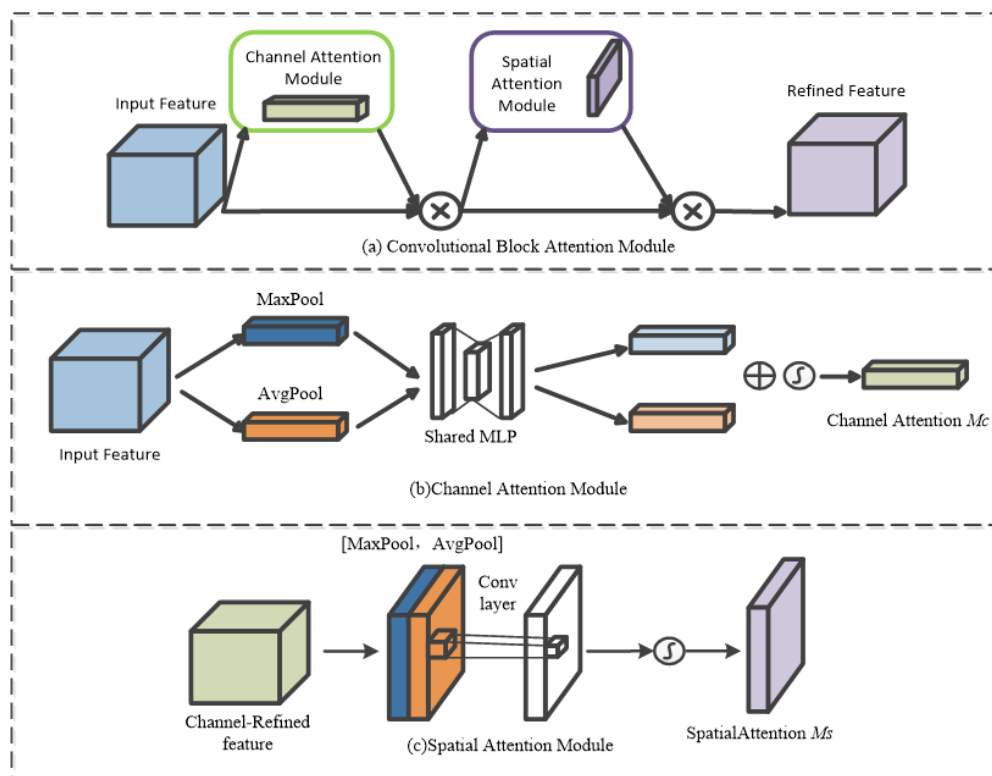


Figure 3. CBAM module structure diagram.

The channel attention mechanism module is introduced to enhance the attention to object features. The calculation is performed as per Equation (1):

$$\begin{aligned} M_c(F) &= \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\ &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \end{aligned} \quad (1)$$

where F is the input feature map, $M_c(F)$ channel attention output weight, W_0 and W_1 are the weight matrix of the first and second fully connected layer, MLP is the shared fully connected layer, σ is the Sigmoid operation, and F_{avg}^c and F_{max}^c are the feature maps after average pooling and maximum pooling on the channel.

By introducing the spatial attention mechanism module, the location information of which is found clustered the most in the direction of the channel as a way of focusing on the location information of the object, the calculation is performed as per Equation (2):

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \\ &= \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \end{aligned} \quad (2)$$

where F is the input feature map, $M_s(F)$ is the output weight of spatial attention, F_{avg}^c and F_{max}^c are the spatial feature map after average pooling and maximum pooling, and $f^{7 \times 7}$ is the convolution operation with a convolution kernel size of 7×7 .

As shown in Figure 4a of the backbone network, the process of downsampling will reduce the model's ability to extract object features, thus reducing the detection accuracy. To solve this problem, this paper embedded the CBAM module in the backbone network and added it after each Conv block from Layer 2 to Layer 4, as shown in Figure 4b. In this way, the detection model can capture key information more efficiently, improve the weight of features related to pests and diseases, improve the effectiveness of features, and improve classification accuracy by establishing interdependencies between convolutional feature channels. In addition, CBAM module can effectively compensate for the small object information lost via multiple convolution operations, and it can improve the sensitivity of the model to small pests and diseases.

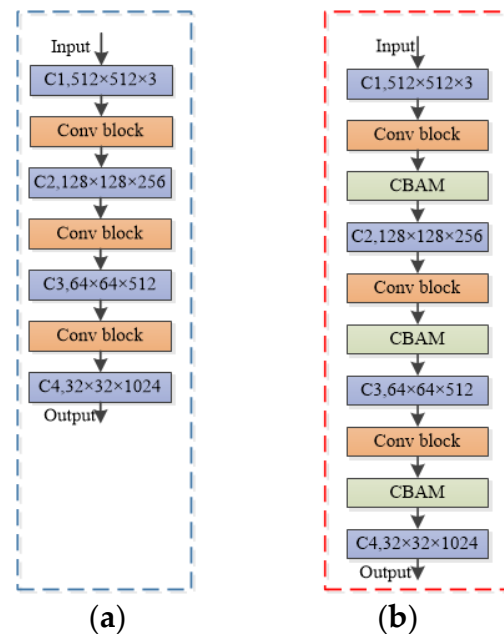


Figure 4. Backbone network structure diagram. (a) Original backbone network; (b) improved backbone network.

2.2.2. Multi-Scale Feature Fusion Module

After many convolution and pooling operations, the features of small-sized pests and diseases become blurred and difficult to extract. Even these features may get deformed and lost. In the object detection task, to improve the detection ability of multi-scale objects, a common method is to fuse the shallow features and deep features together. However, direct fusion of features reduces the capacity of multi-scale representation. Furthermore, the predictions between layers are independent of each other, making it easy to ignore some important features.

To solve these problems, Liu et al. proposed a bidirectional feature fusion network called PANet [13]. PANet achieves bidirectional convergence by introducing bottom-up connectivity. As shown in Figure 5, there are some nodes in PANet with only input edges, which do not contribute much to the performance of the network, and they add additional parameters and computations.

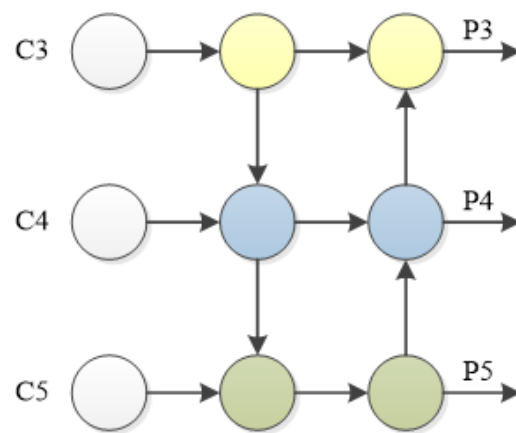


Figure 5. PANet structure diagram.

A weighted bidirectional feature fusion module (MFF) is proposed in this paper. First, the MFF module removes nodes with only input edges and the last node in layer 6. These nodes only pass input features without feature fusion, which contributes little to the network performance and increases the number of parameters. Second, the MFF module adds skip connections to establish connections between inputs and outputs at the same scale layer, both to avoid excessive computational costs and to better fuse features. Since input features with different resolutions contribute differently to the output, additional weights are added for each input feature to learn their importance in the output. A fast normalization fusion method with weights is adopted. Finally, the four feature maps with different resolutions are aggregated into high-resolution output through up-sampling and other operations to summarize multi-scale local information and global information to improve the performance of object detection.

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \bullet I_i \quad (3)$$

where O is the output value of the node, I_i is the input value from node I , w_i and w_j are the input weights of the corresponding nodes. A small quantity ε is guaranteed to be numerically stable, $\varepsilon = 0.0001$.

The structure of the MFF module is shown in Figure 6.

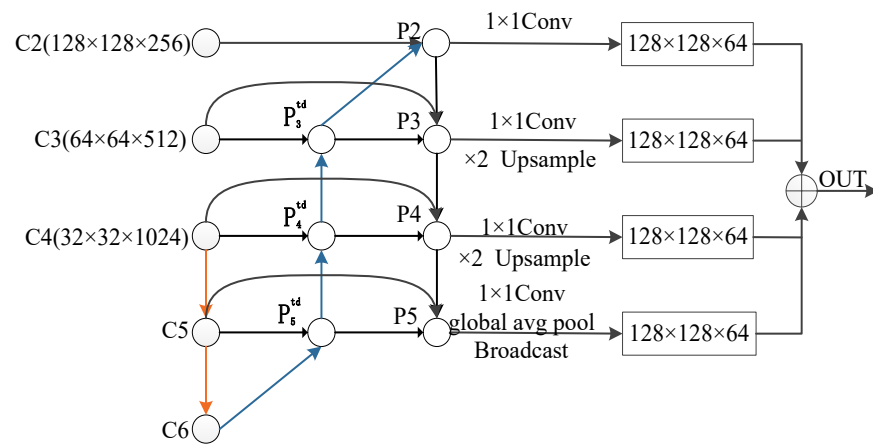


Figure 6. MFF structure diagram.

First, C4 is downsampled to generate C5 and C6. Then, each node is appended with weights for fast normalized feature fusion. Taking the third layer as an example, P_3^{IN} is the input of the third layer, P_3^{td} is the intermediate node of the third layer, and P_3^{out} is the output of the third layer with weight fusion.

The intermediate node features are calculated as shown in Equation (4).

$$P_3^{td} = \text{Conv}\left(\frac{\omega_1 \bullet P_3^{IN} + \omega_2 \bullet \text{Resize}(P_4^{td})}{\omega_1 + \omega_2 + \varepsilon}\right) \quad (4)$$

where Conv is the convolution operation, Resize is the size scaling operation, and ω_1 and ω_2 are the learnable weights of the corresponding node inputs.

The output node features with weight fusion are calculated as in (5).

$$P_3^{out} = \text{Conv}\left(\frac{\omega_1' \bullet P_3^{IN} + \omega_2' \bullet P_3^{td} + \omega_3' \bullet \text{Resize}(P_2^{out})}{\omega_1' + \omega_2' + \omega_3' + \varepsilon}\right) \quad (5)$$

where ω_1' , ω_2' , and ω_3' are the learnable weights entered by the corresponding node.

After fast normalization fusion with weights, four layers of features with different resolutions are obtained. These features are then adjusted to the same resolution and fused into one feature.

3. Experimental Results and Analysis

3.1. Experimental Environment Configuration and Evaluation Metrics

The experimental environment configurations are listed in Table 2.

Table 2. Experimental environment configuration.

Parametric	Configure
Operation System	Ubuntu18.04
CPU	Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50 GHz
GPU	NVIDIA GeForce RTX 2080 Ti
Graphics Card Memory	11 GB
Programming Language	Python3.8
Graphics Acceleration Environment	CUDA 11.3, cuDNN 8
Deep Learning Framework	Pytorch 1.11.0, torchvision 0.12.0

Indices used to evaluate the detection effect of multiple object categories include the mean average precision (mAP) and the average precision (AP) for each category. The AP is determined by the Recall rate (R) and the Precision rate (P), and it is an intuitive standard

for the model performance results of a single category. The F1 score takes into account both the precision and recall of the model.

Recall is calculated using the equation:

$$R = \frac{T_P}{T_P + F_N} \times 100\% \quad (6)$$

Precision is calculated using the equation:

$$P = \frac{T_P}{T_P + F_P} \times 100\% \quad (7)$$

where T_P is the number of correctly detected object objects, F_N is the number of correctly detected object objects, and F_P is the number of incorrectly detected object objects.

The higher the AP and mAP scores, and the higher the FPS, the better the detection performance and speed of the model. The F1 score is also a measure of model performance, taking into account both precision and recall.

3.2. Experimental Results

To evaluate the performance of the proposed method in object detection, the model was trained on the constructed dataset using the Adam optimizer with iterative training. The initial learning rate was 5×10^{-4} , and the learning rate was adjusted by cosine annealing during training. The weights of the CenterNet-ResNet50 network pre-trained on Imagenet were used to initialize the network parameters, and the entire network was trained for 200 epochs with transfer learning. The loss curves are shown in Figure 7. It can be seen that the loss value decreased rapidly in the first ten cycles and then became basically stable after 150 cycles of training, indicating that the model gradually learned the object detection features.

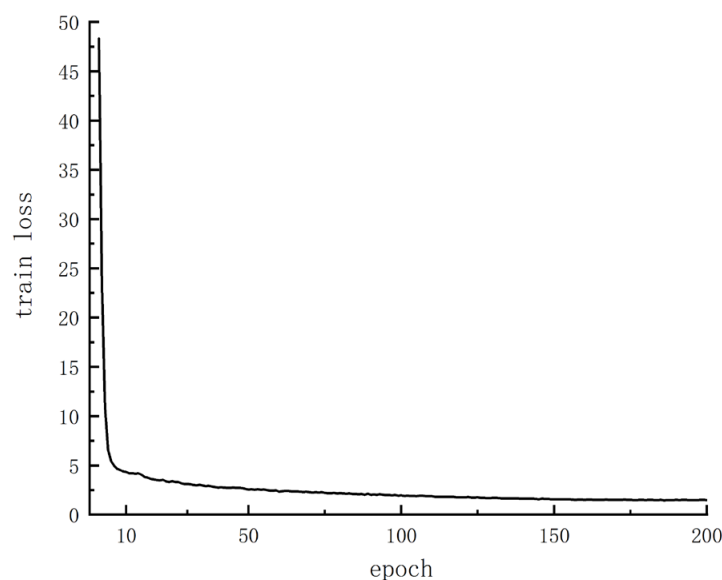


Figure 7. Model training loss curve.

In the experiments of this paper, the test results of each class were evaluated, and the accuracy and average test time were calculated. According to the results shown in Table 3, this research model demonstrates good performance in the detection of different object categories. Taking maize borer infestation as an example, the accuracy of the model reached 93.98%, which proves that the model can accurately identify maize borer infestation. For northern leaf blight (NLB) disease, a detection accuracy of 86.60% was obtained, indicating that the model is highly reliable in detecting this type of disease.

Table 3. Model detection results in this paper.

Label Name	AP (%)	P (%)	R (%)	F1 Score
Maize borer	93.98	94.44	72.34	0.82
Army worm	91.50	92.31	75.00	0.83
NLB	86.60	89.55	64.52	0.75
Red spider	88.27	93.75	83.33	0.88
Peach borer	81.90	88.89	69.57	0.78
Aphids	87.69	90.60	67.95	0.78
NLS	76.35	96.25	24.06	0.39
GLS	74.73	93.75	25.62	0.40

Moreover, in terms of average detection time, the model in this paper takes only 0.025 s, indicating the fast processing capability of the model while efficiently detecting the object. Therefore, the model is able to detect both diseases and pests with good results in terms of speed and accuracy.

3.3. Analysis of Experimental Results

3.3.1. Ablation Experiment

To verify the performance of the attention module proposed in this paper, we selected and compared the widely used SENet [14], ECA-Net [15], and SimAM [16] attention modules with the CBAM module used in this paper. By embedding them into the CenterNet backbone network, we obtained four models: CenterNet+SENet, CenterNet+ECA-Net, CenterNet+SimAM, and CenterNet+CBAM.

As can be seen from the results in Table 4, the model in this paper has a higher recall and can detect more pests and diseases than the other attention modules. The average F1 score is also higher than that of the SENet, ECA-Net, and SimAM models, with improvements of 0.11, 0.10, and 0.09, respectively. The CBAM module focuses on both channel and spatial attention mechanisms, while SENet and ECA-Net focus on channel direction only, and SimAM is a 3D attention module that differs from existing channel and spatial attention modules.

Table 4. Comparison of results by different attention mechanisms.

Model	P (%)	R (%)	F1 Score	mAP (%)
CenterNet+SENet	96.17	43.39	0.58	78.27
CenterNet+ECA-Net	94.87	44.71	0.59	76.97
CenterNet+SimAM	96.98	48.10	0.60	77.51
CenterNet+CBAM	91.39	61.03	0.69	78.62

Compared with the CenterNet+SENet, CenterNet+SimAM and CenterNet+ECA-Net models, the proposed model improves the mAP by 0.44%, 1.41%, and 2.1%, respectively. This indicates that the CBAM module chosen in this paper can effectively calibrate the channel and spatial features, thus improving the detection performance of the algorithm.

To further demonstrate the effectiveness of the modified module in the proposed algorithm, ablation experiments were conducted to compare the performance of the following four sets of models:

- (1) In the existing CenterNet model, the backbone network is ResNet50, and the neck network is the deconvolution network layer, hereinafter referred to as ResNet50+deconvolution layer;
- (2) The CBAM module is embedded in the backbone network of the original CenterNet model, hereinafter referred to as ResNet50+CBAM+deconvolution layer;
- (3) The backbone network is ResNet50, the new MFF module is added as the neck network, and the deconvolution layer of the original CenterNet model is deleted, hereinafter referred to as ResNet50+MFF;

- (4) The algorithm in this paper is based on (3) to embed the CBAM module in the backbone network, hereinafter referred to as ResNet50+CBAM+MFF.

As can be seen from the results in Table 5, the addition of CBAM module and MFF module significantly improves the model accuracy. Especially after adding the MFF module, the mAP of the ResNet50+MFF model improved by 6.96% compared to the original model. Moreover, the mAP of the ResNet50+CBAM+deconvolutional layer model with CBAM module embedded in the backbone network is 3.08 percent higher than that of the original model. This indicates that these two separate modules are very helpful for model detection.

Table 5. Ablation experiments based on improved CenterNet model.

Model	P (%)	R (%)	F1 Score	mAP (%)	Parameters (M)
ResNet50+Deconvolutional (Original)	95.36	32.01	0.45	75.54	32.665
ResNet50+CBAM+Deconvolutional	91.39	61.03	0.69	78.62	33.714
ResNet50+MFF	91.21	57.97	0.68	82.50	23.952
ResNet50+CBAM+MFF (Ours)	92.11	60.30	0.70	85.13	24.296

When the deconvolutional layer of the original model is removed after adding the MFF module, the number of parameters in the ResNet50+MFF model is reduced by 26.67%. Compared to the deconvolutional layer of the original model, the MFF module has a smaller parameter scale, and the number of parameters of the model is reduced, while the accuracy of the model is greatly improved. By adding these two modules for training at the same time, the model in this paper (ResNet50+CBAM+MFF) reaches the highest mAP, which is 9.59% higher than the original model. The improved model also achieved the highest F1 score of 0.70, indicating that the integrated performance of the model is also better.

In summary, the model in this paper significantly improves the accuracy while reducing the number of parameters, demonstrating that the attention mechanism and multi-scale feature fusion module (MFF) can significantly enhance the comprehensive performance of the model, and is more suitable for the detection of maize pests and diseases.

3.3.2. Comparative Analysis Experiment

The detection results of the proposed model and the original CenterNet model on the test set are compared and analyzed. In the detection results, all detected objects are marked with a border, and the label name and confidence level of the object are displayed. It can be seen from Figure 8 that the original model misses detection when the infestation is clustered, and from Figure 9, it can be seen that the original model misdetects the GLS as the NLS at the same time. The improved model uses an attention mechanism to suppress interfering objects, integrates multi-scale features, improves the detection performance of objects, effectively detects object outcomes, and has higher confidence in prediction results.

To further evaluate the performance of this model, a comparison experiment was conducted using mainstream object detection algorithms. The dataset used in the experiments was the maize leaf pests and diseases dataset from this paper, and the same experimental parameters and experimental environment were set. Disease and insect detection were performed between the proposed model and the Faster R-CNN [17], YOLOv5, SSD-VGG [18], and Efficientdet-D0 [19] models. Table 6 and Figure 10 show that compared with SSD-VGG, YOLOv5, Faster R-CNN, and Efficientdet-D0 methods, the mean average precision (mAP) of the proposed model was the highest among the five detection algorithms, increasing by 9.17%, 7.34%, 4.43%, and 2.46%, respectively. This demonstrates the effectiveness of the proposed method.

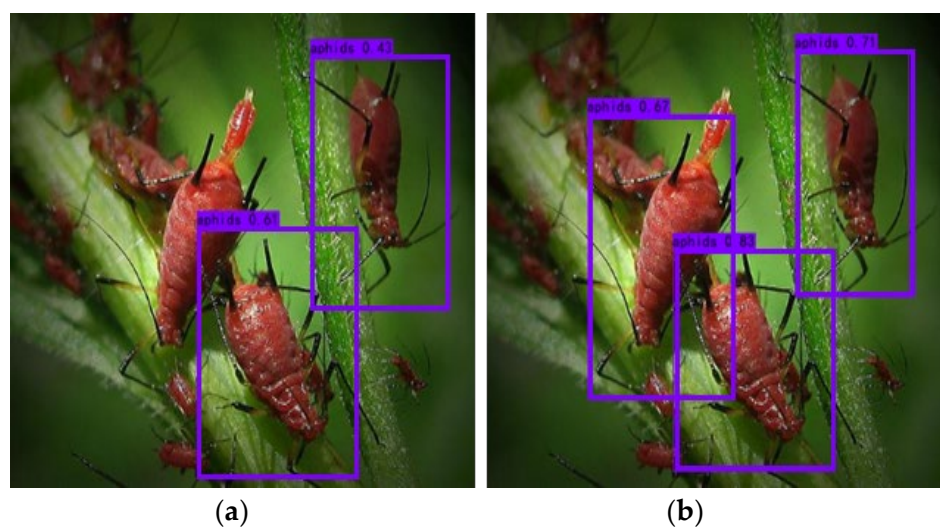


Figure 8. Comparison of detection results for leakage cases. (a) Original CenterNet model; (b) our model.

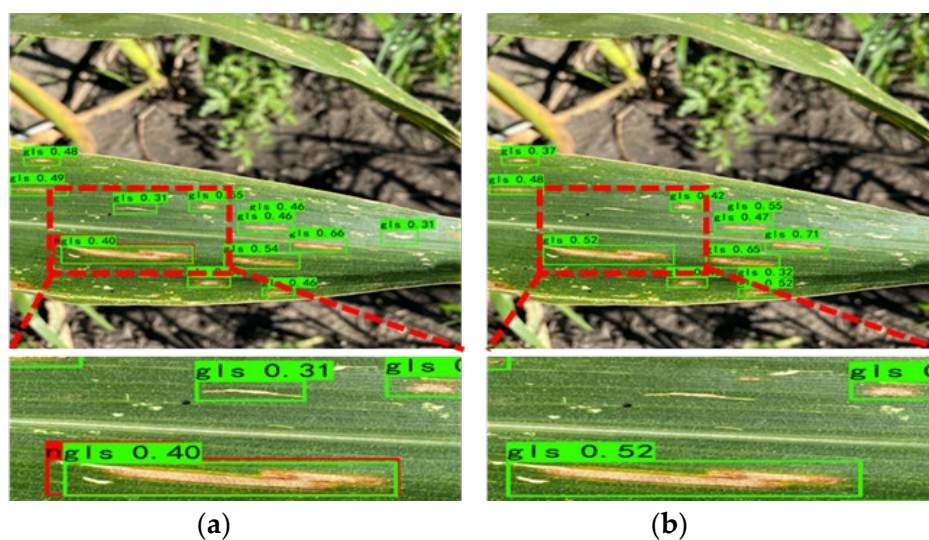


Figure 9. Comparison of detection results in case of misdetection. (a) Original CenterNet model; (b) our model.

Table 6. Pest detection results of different object detection models.

Model	P (%)	R (%)	F1 Score	mAP (%)	Parameters (M)	FPS (f/s)
SSD-VGG	78.63	65.82	0.68	74.17	26.285	58.13
CenterNet	95.36	32.01	0.45	75.54	32.665	57.60
Faster R-CNN	54.10	82.55	0.64	76.00	137.099	12.29
YOLOv5	90.83	45.98	0.59	78.91	47.057	48.18
Efficientdet-D0	87.23	71.62	0.76	82.67	3.874	14.63
Ours	92.11	60.30	0.70	85.13	24.296	23.69

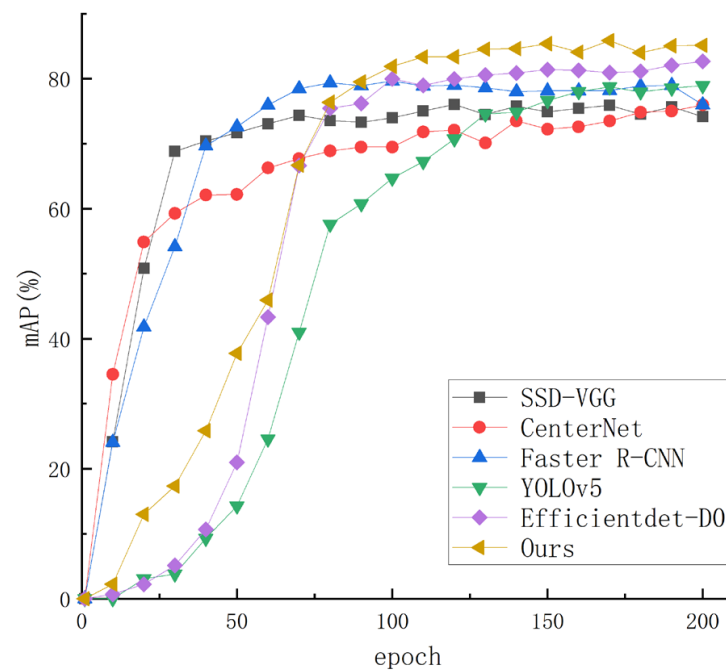


Figure 10. mAP curves for six different object detection models.

The SSD-VGG algorithm is a one-stage algorithm. It first performs dense sampling uniformly at different locations on the image and then uses convolutional neural networks for feature extraction, followed by classification and regression. The SSD-VGG algorithm has the advantage of its fast speed, which is the fastest among the six algorithms, but its detection accuracy is insufficient. Compared to the SSD-VGG algorithm, Faster R-CNN has a higher detection accuracy, but the parameter scale is large due to its two-stage algorithm. The YOLOv5 algorithm improves the detection accuracy and parameter scaling to some extent, but its detection performance is poor on this dataset. The Efficientdet-D0 algorithm has the smallest parameter size among the six algorithms, but it uses more convolution operations, resulting in slower model detection.

In order to visually demonstrate the detection performance of the proposed algorithm, the detection results of the proposed algorithm in the test set were compared and analyzed with those of Faster R-CNN, YOLOv5, SSD-VGG, and Efficientdet-D0 algorithms. The detection results were presented in Figure 11. As shown in Figure 11, the first column represents images with GLS disease, the second column represents images with Aphids pest, and the third column represents images with NLB disease. It can be observed from Figure 11 that while Faster R-CNN detected a larger number of targets, it also had a significant number of false detections. On the other hand, YOLOv5, SSD-VGG, and Efficientdet-D0 displayed missing detections in certain instances. Additionally, the Efficientdet-D0 algorithm misclassified GLS disease as NLB disease in the first column. In the third column, both Faster R-CNN and YOLOv5 were able to detect the fuzzy small target, while the other three algorithms failed to detect it. Overall, in all three columns of images except for the fuzzy small target in the third column, all other targets of pests and diseases were successfully detected. Based on this analysis, it can be concluded that compared with the comparison algorithm, the proposed algorithm is more accurate in the detection of pests and diseases.

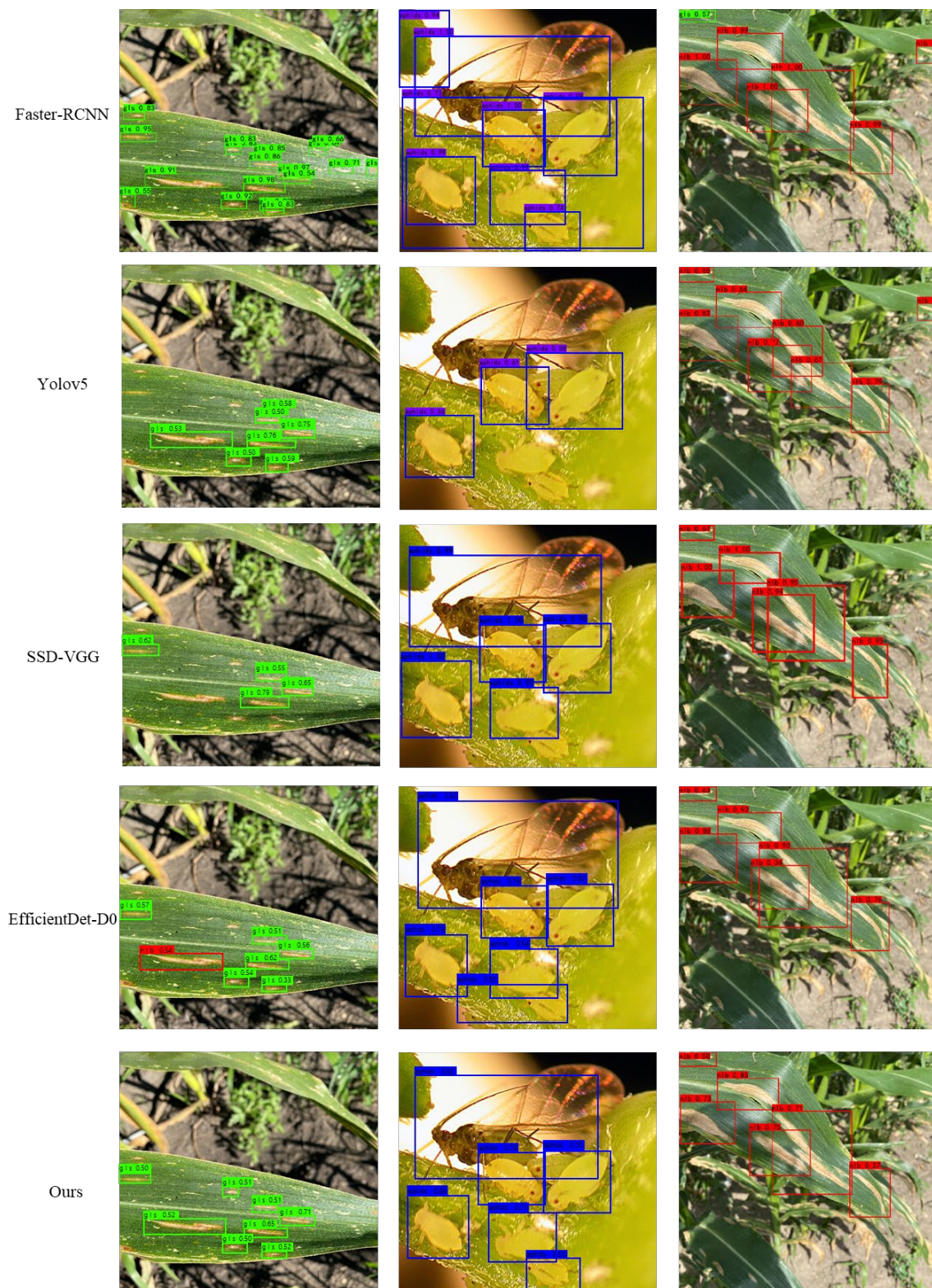


Figure 11. Comparison of the visualization of the detection results of different model.

3.4. Web Detection Platform

To facilitate farmers in easily and quickly identifying and detecting pests and diseases, the model is deployed in a web application that detects pests and diseases on maize leaves, assisting farmers in understanding the conditions of pests and diseases and obtaining expert guidance. In this application, users can detect maize leaf pests and diseases by

uploading local pictures, and the test results are presented in Figure 12. Additionally, the application utilizes an embedded database, SQLite, to construct a web user service system, including registration and login functions, maize leaf pest detection functions, encyclopedia search functions, and pest control advice functions. Through the above functional design and implementation, the application provides users with convenient services and practical functions.

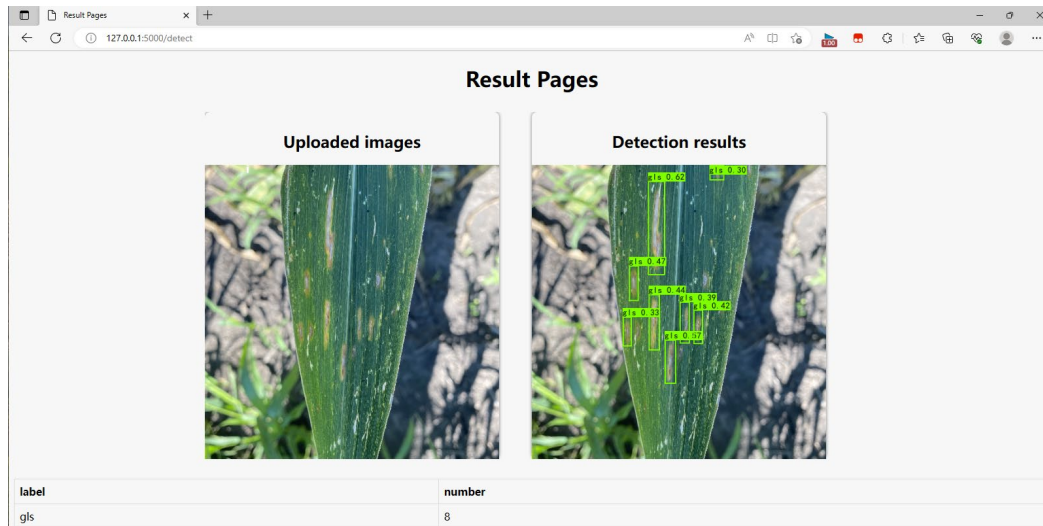


Figure 12. Web application interface diagram.

4. Discussion

Detecting leaf pests and diseases in agriculture is critical to ensuring crop health and yield. In recent years, various machine learning and deep learning models have been developed to automate this process, leading to more efficient and accurate detection. In this case, we evaluated the performance of our proposed model compared to existing state-of-the-art methods. We evaluated their model's accuracy, parameter size, detection speed, and dataset to understand its strengths and limitations.

Based on Table 7, several observations can be made. When it comes to accuracy, our model, with an mAP of 85.13%, represents a significant improvement over the Pest R-CNN model by Du et al., 2022 [20], which achieved an mAP of 60.2%. However, when it comes to maize disease detection, the EfficientNet model from Liu et al., 2022 [21] slightly outperforms ours with an accuracy of 98.52%. Zhang et al., 2020 [22] effentnet-b4 model accuracy reaches 97%, but the model size is larger, 268.62 M, while our model not only has a smaller parameter size, 24.296 M, but also realizes a competitive mAP, highlighting its efficiency. In terms of detection speed, our model can process each image in a mere 0.025 s, outpacing the ResNet50 model from Shin et al., 2021 [23], which takes 0.076 s per image. This rapid detection speed is crucial for real-time applications in agriculture. The diversity and quality of datasets used for training play a crucial role in model performance, with many studies like those by Zeng et al., 2022 [24], and Chen et al., 2021 [25], utilizing the PlantVillage dataset. However, our model is trained on a dataset collected from the network, which may contain a wider range of scenarios and conditions than the PlantVillage dataset. In essence, while each model has its own strengths and weaknesses, our proposed model strikes a commendable balance between accuracy, efficiency, and speed. The ability to detect both disease and pest tasks further increases its potential for practical agricultural applications.

Table 7. Comparison of proposed advanced pest and disease detection method with other state-of-the-art approaches on maize and other plant datasets.

References	Type of Plant	Dataset	Technique	Performance
(Zhang et al., 2020) [22]	Cucumber	Private	EfficientNet-B4	Accuracy: 97% Parameters: 268.62 M
(Shin et al., 2021) [23]	Strawberry	Private	ResNet50	Accuracy: 98.11% Detect each image: 0.076 s
(Jiang et al., 2021) [26]	Paddy Wheat	Dataset provided by website	VGG16	Paddy: 97.22% Wheat: 98.75 Parameters: 68 M
(Elfatimi et al., 2022) [27]	Beans	public dataset presented by tensorflow	MobileNetV2	Accuracy: 92.0%
(Zeng et al., 2022) [24]	Maize (disease)	PlantVillage, Private	Lightweight dense-scale network	Accuracy: 95.4% Parameters: 0.59
(Chen et al., 2021) [25]	Maize (disease)	PlantVillage and private	DenseNet	Accuracy: 95.86%
(Liu et al., 2022) [21]	Maize (disease)	Crop disease AI Challenge dataset	EfficientNet	Accuracy: 98.52% Parameters: 46.21 M
(Yin et al., 2022) [28]	Maize (disease)	Private	DISE-Net	Accuracy: 97.12%
(Du et al., 2022) [20]	Maize (Pest)	private	Pest R-CNN	mAP: 60.2% mAP: 85.13%
Ours	Maize (disease and pest)	Dataset provided by website	Enhancement of CenterNet.	parameters: 24.296 M Detect each image: 0.025 s

Future research can be approached from several perspectives. Firstly, the model's training can be deepened further, and various data enhancement techniques can be employed to enhance the model's generalization ability. Secondly, research can be combined with other fields, such as remote sensing technology, to achieve rapid detection of pests and diseases on a larger scale. Finally, lightweight optimization of the model needs to be considered to enable it to run on low-performance equipment, thereby extending its practical application in the field.

5. Conclusions

In this paper, we construct a model for corn leaf disease and insect detection that integrates multi-scale features and attention mechanisms to detect three corn leaf diseases and five pests. The CBAM attention module is added to improve the ability to extract relevant features of pests and diseases. The MFF module is designed in the neck network, and a weighted bidirectional feature fusion network is employed to enable the model to better detect multi-scale objects. The experimental results demonstrate that the constructed model achieves 85.13% accuracy on the test dataset, which is 9.59% higher than the accuracy of the original CenterNet model. The improved model has a parameter scale of 24.296 M, which is 25.7% smaller than the original CenterNet model. It takes 0.025 s to predict a 512×512 pixel image, and the processing speed reaches 23.69 f/s, which can meet the requirements of real-time detection. Compared with Faster R-CNN, YOLOv5, SSD-VGG, and Efficientdet-D0 models, the proposed model achieves better detection accuracy under the same experimental conditions. Therefore, this model can achieve rapid and accurate detection of pests and diseases and meet the requirements of pest and disease detection. The model is deployed in a web application that allows users to upload photos for quick detection, providing users with convenient services and practical features to assist farmers in understanding the situation of pests and diseases. In the future, a lightweight model will be designed, and the model structure will be further optimized using distillation technology.

Author Contributions: Investigation, J.K.; methodology, W.Z.; software, W.Z.; supervision, J.K.; writing—original draft preparation, W.Z.; writing—review and editing, J.K., W.L. and Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Youth Fund of the National Natural Science Foundation of China (Grant No. 62203285) and the Youth Fund of Shaanxi Provincial Natural Science Basic Research Program General Project (Grant No. 2022JQ-181) and was also funded by the Xi'an Science and Technology Plan Project (Grant No. 23NYGG0070).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhai, Z.Y.; Cao, Y.F.; Xu, H.L.; Yuan, P.S.; Wang, H.Y. Review of key techniques for crop disease and pest detection. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 1–18.
2. Shao, M.Y.; Zhang, J.H.; Feng, Q.; Chai, X.J.; Zhang, N.; Zhang, W.R. Research progress of deep learning in detection and recognition of plant leaf diseases. *Smart Agric.* **2022**, *4*, 29–46.
3. Ahmad, A.; Saraswat, D.; Gamal, A.E.; Gurmukh, J. CD&S Dataset: Handheld imagery dataset acquired under field conditions for maize disease identification and severity estimation. *arXiv* **2021**, arXiv:2110.12084.
4. Wang, Z.Y.; Wang, X.M. Current status and management strategies for maize pests and diseases in China. *Plant Prot.* **2019**, *45*, 1–11.
5. Guo, X.Y.; Yu, S.Q.; Shen, H.C.; Li, L.; Du, J.J. Deep Learning Network for Crop Disease Recognition with Global Feature Extraction. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 301–307.
6. Xie, X.; Ma, Y.; Liu, B.; He, J.; Li, S.; Wang, H. A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Front. Plant Sci.* **2020**, *11*, 751. [[CrossRef](#)] [[PubMed](#)]
7. Liu, J.; Wang, X. Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. *Front. Plant Sci.* **2020**, *11*, 898. [[CrossRef](#)]
8. Richey, B.; Shirvaikar, M.V. Deep learning based real-time detection of Northern Maize Leaf Blight crop disease using YoloV4. In *Proceedings of the Defense + Commercial Sensing; SPIE: Bellingham, WA, USA*, 2021.
9. Sun, H.N.; Xu, H.W.; Liu, B.; He, D.J.; He, J.R.; Zhang, H.X.; Geng, N. MEAN-SSD: A novel real-time detector for apple leaf diseases using improved light-weight convolutional neural networks. *Comput. Electron. Agric.* **2021**, *189*, 106379. [[CrossRef](#)]
10. Yang, S.Q.; Liu, J.C.; Xu, K.K.; Sang, X.; Ning, J.F.; Zhang, Z.T. Improved CenterNet based maize tassel recognition for UAV remote sensing image. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 206–212.
11. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.
12. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module; Springer: Cham, Switzerland, 2018; pp. 3–19.
13. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–23 June 2018.
14. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–23 June 2018.
15. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13–19 June 2020.
16. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In *Proceedings of the International Conference on Machine Learning*, Virtual, 18–24 July 2021.
17. 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.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
18. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Proceedings of the Computer Vision—ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
19. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13–19 June 2020.
20. Du, L.; Sun, Y.; Chen, S.; Feng, J.; Zhao, Y.; Yan, Z.; Zhang, X.; Bian, Y. A Novel Object Detection Model Based on Faster R-CNN for *Spodoptera frugiperda* According to Feeding Trace of Maize Leaves. *Agriculture* **2022**, *12*, 248. [[CrossRef](#)]
21. Liu, J.; Wang, M.; Bao, L.; Li, X. EfficientNet based recognition of maize diseases by leaf image classification. *J. Phys. Conf. Ser.* **2020**, *1693*, 012148. [[CrossRef](#)]

22. Zhang, P.; Yang, L.; Li, D. EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment. *Comput. Electron. Agric.* **2020**, *176*, 105652. [[CrossRef](#)]
23. Shin, J.; Chang, Y.K.; Heung, B.; Nguyen-Quang, T.; Price, G.W.; Al-Mallahi, A. A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Comput. Electron. Agric.* **2021**, *183*, 106042. [[CrossRef](#)]
24. Zeng, W.; Li, H.; Hu, G.; Liang, D. Lightweight dense-scale network (LDSNet) for maize leaf disease identification. *Comput. Electron. Agric.* **2022**, *197*, 106943. [[CrossRef](#)]
25. Chen, J.; Wang, W.; Zhang, D.; Zeb, A.; Nanekaran, Y.A. Attention embedded lightweight network for maize disease recognition. *Plant Pathol.* **2021**, *70*, 630–642. [[CrossRef](#)]
26. Jiang, Z.; Dong, Z.; Jiang, W.; Yang, Y. Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. *Comput. Electron. Agric.* **2021**, *186*, 106184. [[CrossRef](#)]
27. Elfatimi, E.; Eryigit, R.; Elfatimi, L. Beans Leaf Diseases Classification Using MobileNet Models. *IEEE Access* **2022**, *10*, 9471–9482. [[CrossRef](#)]
28. Yin, C.; Zeng, T.; Zhang, H.; Fu, W.; Wang, L.; Yao, S. Maize Small Leaf Spot Classification Based on Improved Deep Convolutional Neural Networks with a Multi-Scale Attention Mechanism. *Agronomy* **2022**, *12*, 906. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.