applied sciences

MDPI

Article

Research on Intelligent Recognition for Plant Pests and Diseases Based on Improved YOLOv8 Model

Yuchun Wang [1], Cancan Yi [2,3,4,*], Tao Huang [2,3,4] and Jun Liu [2,3,4]

[1] School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China; 17386375361@163.com
[2] Key Laboratory of Metallurgical Equipment and Control Technology, Wuhan University of Science and Technology, Ministry of Education, Wuhan 430081, China; huangtao18@wust.edu.cn (T.H.); liujunesix@163.com (J.L.)
[3] Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China
[4] Precision Manufacturing Institute, Wuhan University of Science and Technology, Wuhan 430081, China
* Correspondence: meyicancan@wust.edu.cn

Abstract: Plant pests and diseases are important parts of insect disease control and the high-quality development of agriculture. Traditional methods for identifying plant diseases and pests suffer from low accuracy and slow speed, while the existing machine learning methods are constrained by environmental and technological factors, leading to low recognition efficiency. To address the issue of the above problems, this paper has proposed an intelligent recognition algorithm based on the improved YOLOv8 model, which has high recognition accuracy and speed. Firstly, in the Backbone network, the Global Attention Mechanism (GAM) is adopted to weigh the important feature information, thereby improving the accuracy of the model. Secondly, in the mixed feature part of the Neck network, the Receptive-Field Attention Convolutional (RFA Conv) operation is used instead of standard convolution operations to enhance the processing ability for feature information and to reduce computational complexity and costs, thus improving the network performance. After verifying the rice and cotton datasets, the accuracy indicator mean average precision (mAP) reaches 71.27% and 82.91%, respectively, in the two different datasets. Comparing these indices with those of the Faster R-CNN, YOLOv7, and the original YOLOv8 model, the results fully demonstrate the effectiveness and superiority of the improved model in terms of detection accuracy.

Keywords: plant pests and diseases recognition; YOLOv8 model; GAM; RFA Conv; disease identification

check for updates

Citation: Wang, Y.; Yi, C.; Huang, T.; Liu, J. Research on Intelligent Recognition for Plant Pests and Diseases Based on Improved YOLOv8 Model. *Appl. Sci.* **2024**, *14*, 5353. https://doi.org/10.3390/app14125353

Academic Editor: Christos Bouras

Received: 9 May 2024
Revised: 11 June 2024
Accepted: 18 June 2024
Published: 20 June 2024

Copyright: © 2024 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

Agriculture is a primary industry that supports the national economy and livelihoods of people. The yield and quantity of crops are closely linked to the high-quality development of agriculture. However, recent studies reveal that crop diseases and pests are on the rise due to factors such as global climate change [1], changes in farming practices [2], and excessive use of pesticides. These factors have a significant impact on agricultural production and development [3]. Early identification of crop pests and diseases can help determine their categories and range of spread, facilitating early monitoring and control. In addition, improving crop pest and disease monitoring and control systems helps to effectively control the spread [4]. Therefore, the identification of crop pests and diseases becomes a key link in the prevention and control system, and improving the efficiency and accuracy of identifying crop pests and diseases is of great significance in reducing agricultural production losses and promoting high-quality agricultural development.

The conventional method for the identification of crop pests and diseases relies on manual recognition that is based on the judgment and experience of professional researchers and farmers, which is slow and has low accuracy [5]. With the development of image

processing technology, the recognition method of traditional machine learning, such as Image Processing Techniques (IPT) and Machine Learning Algorithms (MLA), is used [6], which greatly improves efficiency compared to manual recognition. Nevertheless, this method is limited by factors such as the target position and the lighting background. Furthermore, when trying to use human-made feature extraction and identifying plant diseases, it is difficult to predict which combination of pre-processing, feature extraction, and classification algorithm will produce the best results, leading to tedious trial and error. Therefore, this method cannot meet the demands of complex real-world environments in terms of accuracy and speed, giving it relatively low applicability.

With the rapid development of deep learning and the sharp improvement of computing power, the target detection algorithm has shifted from traditional algorithms based on manual characteristics to deep learning-based target detection techniques. The target detection algorithm based on deep learning uses the Convolutional Neural Network (CNN) to extract features. Through model training and parameter optimization of a large amount of image data, the target algorithm enables the automatic discovery of the necessary feature information for detecting and classifying targets, meeting the requirements for speed and accuracy in target detection. Target detection algorithms based on deep learning can be divided into two categories according to the algorithmic process: Two-stage target detection algorithms and One-stage target detection algorithms. The Two-stage detection algorithms represented by the R-CNN series use the Selective Search method to generate several candidate areas for an image, utilize the CNN to achieve extraction features for each candidate area, and finally make category judgments and adjust the selection frame positions [7]. The One-stage target detection algorithms represented by the YOLO series [8] and SSD series [9] use neural networks to process images and achieve direct prediction of target categories and positions through end-to-end detection methods, thereby improving computation speed. The above two target detection algorithms are applied in the field of recognition of plant pests and diseases. Lin Jiao et al. [10] put forward an Anchor-Free Region Convolutional Neural Network (AF-RCNN) based on fused feature maps, integrating it with the Fast R-CNN into a single network to detect 24 classes of pests via an end-to-end method. The proposed method exhibits certain real-time performance and accuracy. Shamse et al. [11] use a method that detects diseases from plant leaf images by Tensorflow, which is an object detection API, and further improve the recognition accuracy of leaves by training the model using the Faster R-CNN method. Yunong Tian et al. [12] add the Densenet module and Adaptive Attention Module (AAM) to the feature extraction part based on YOLOv3. This innovative approach, known as MD-YOLO network, improves the detection accuracy for small-sized target pests. Junzhen Yu, et al. [13] propose a Stolon-YOLO for visual recognition of stolon of strawberry seedlings in glass greenhouses combining two modules: HorBlock-decoupled head and Stem Block feature enhancement module. Compared to YOLOX, it increases the recall rate by 8.3% and increases the accuracy by 12.2%. Liu J et al. [14] optimize the feature layer of the YOLOv3 model by using an image pyramid to achieve multi-scale feature detection, thereby improving the detection accuracy and speed of the Yolov3 model. V Senthil Kumar et al. [15] present a multi-scale YOLOv5 detection network for the early detection and classification of rice crop diseases. The proposed Bidirectional Feature Attention Pyramid Network (Bi-FAPN) is used to extract the features from the segmented image and enhance the detection accuracy for diseases with different scales. D. Luo et al. [16] put forward the Lightweight Self-Attention YOLOv8 model. They introduce an innovative feature fusion technique known as the asymptotic characteristic pyramid network (AFPN) at the Neck, thereby exhibiting an average increase in detection precision of 2.8% compared to YOLOv8. Y. Yang et al. [17] construct the self-made RiPest rice pest dataset and propose an improved YOLOv8 model, named Gi-YOLOv8, which has a 1.3% improvement of accuracy compared to the original YOLOv8 model. Y. Di et al. [18] present a lightweight attention-based network known as TP-YOLO. It introduces Contextual Transformer and Omni-Dimensional Dynamic Convolution modules. These two attention-based components can enhance feature extraction. Uddin MS et al. [19] optimize

YOLOv8s configuration by adding three extra Convolution blocks and using the Swish as the activation function, demonstrating performance in cauliflower disease detection. The methods mentioned above show good application effects in the identification of plant pests and diseases, but there is still room for improvement in the detection accuracy and detection speed. Real-time detection tasks generally require high detection speed. Therefore, this paper conducts research based on a One-stage target detection algorithm known for fast detection speed.

Since its inception, the One-stage target detection algorithm YOLO has received considerable attention and has been widely applied. In recent years, the YOLO series of algorithms have been optimized, with the Ultralytics team introducing the YOLOv8 version in January 2023 [20], which achieves a good balance between accuracy and speed and is suitable for plant pest and disease recognition detection. However, there are certain issues when applying the YOLOv8 model to plant pest and disease recognition detection. Firstly, there is a background characteristic, except for in pest defects in plant leaf images, which will cause interference to detection and reduce both accuracy and speed. Secondly, pest and disease defects on plant leaves often have irregular shapes. The YOLOv8 predicts targets using the centerline, width, and height of predicted bounding boxes. In cases where defects are irregular or closely spaced, the accuracy of predictions will decrease. Thirdly, there are many plant species as well as pest and disease categories, leading to an uneven distribution of sample information and lower accuracy for the samples with small size. Therefore, to address the above issues, this paper optimizes the YOLOv8 model. Firstly, the Global Attention Mechanism (GAM) [21] is introduced to enhance the interaction between channels and spatial dimensions by incorporating a mechanism to preserve information on top of channel and spatial attention mechanisms, leading to improved accuracy. Secondly, the Receptive-Field Attention Convolutional (RFA Conv) [22] method is used to replace standard convolution operations in the current neural networks, which significantly improves network performance without increasing computing costs. Finally, in terms of datasets, considering the differences and similarities in the presence of pests and diseases among different plants, a dataset is created in which rice and cotton are representatives for plants. It includes diseases like Blast, Blight, and Brownspot for rice [23] and Alternaria Leaf Spot, Curl Leaves, Foliar Disease, and Herbicide [24] for cotton, enriching the sample information and improving its applicability.

The main contributions of this paper are summarized as follows:

(1) Add the GAM to improve the model accuracy without adding a calculational burden;
(2) Use the RFA Conv to replace the standard convolution operation, reduce the waste of computational resources, and improve the accuracy;
(3) Make representative datasets containing various pests and diseases to improve applicability;
(4) Use the improved YOLOv8 model for practical application, apply it to the self-made datasets, and compare it to similar algorithms. The results show that the method of this paper yields the best performance in detection accuracy.

2. Related Work

In the One-stage target detection algorithms, the series of YOLO has undergone multiple updates and upgrades. Due to its high accuracy, fast speed, and convenient deployment, it has a wide range of applications. Therefore, the YOLO series was chosen for plant-diseased pest identification. The basic method of YOLO is to adjust the input image to a fixed size, then use the CNN structure to extract features, and finally process the network and output detection results. The early version of YOLO selects partition detection methods instead of sliding windows and classifiers to simplify the operation process. Firstly, divide the input image into a grid of $n \times n$. Then for each grid, the center point of the target being detected must fall within it. Each grid can predict multiple target boxes and their confidence. The output information of each target box includes location information and target categories. Finally, the detection results can be output by non-maximum suppression. Compared to the two outputs of Fast R-CNN probabilistic classification and the bounding

box regression, YOLO reframes object detection as a single regression problem to optimize. The subsequent improvement iterations are based on this to continuously improve the accuracy and speed of detection. At present, YOLOv8 has better performance in the YOLO series, so this paper uses YOLOv8 as the basic model. As a model with good comprehensive performance in the YOLO series, YOLOv8 effectively balances the accuracy and speed of the model as displayed in Figure 1. The network structure of YOLOv8 mainly consists of four parts: Input for input, Backbone for backbone network, Neck for mixed feature, and Head for prediction output.

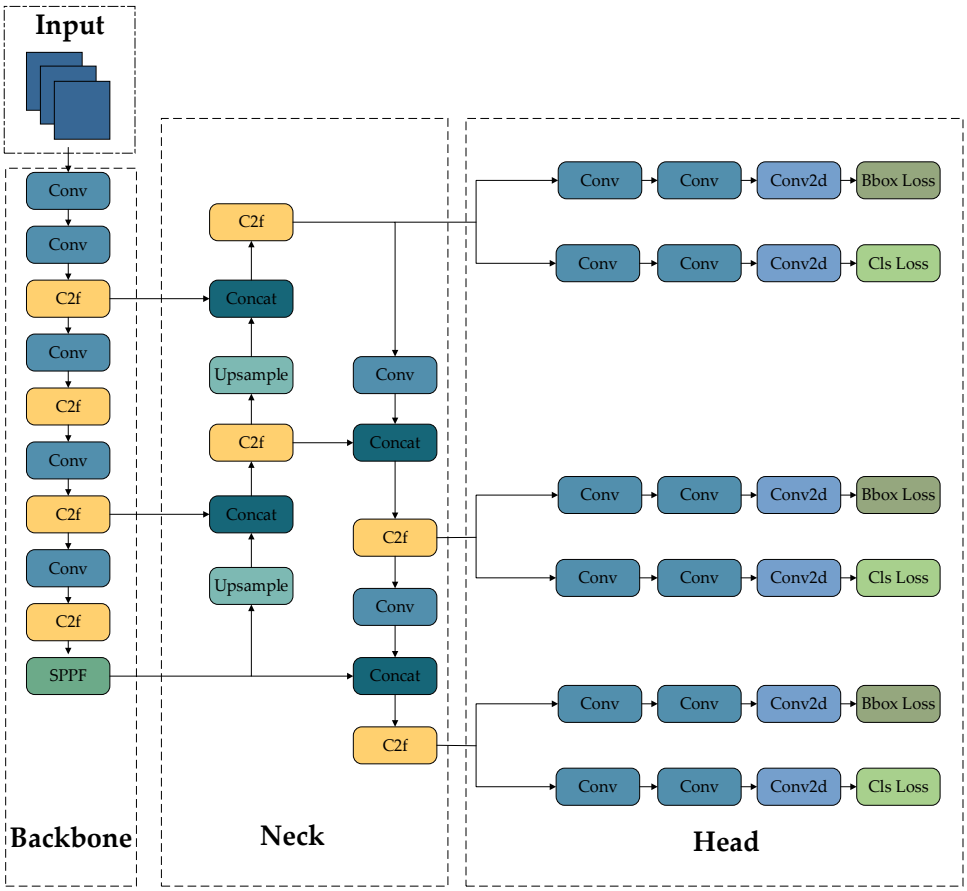


Figure 1. YOLOv8 network structure diagram.

In the first part of the input terminal, the input image is reorganized and defined as a $640 \times 640 \times 3$ three-channel picture. The second part, the Backbone network, includes CBS modules, C2f modules, and Fast Spatial Pyramid Pooling (SPPF) modules, which extract multi-scale features from the input pictures. The SPPF module is a spatial pyramid pooling layer as shown in Figure 2 that can expand the acceptance region, realize local and global characteristics fusion, and enrich characteristic information. Compared to SPP, this model reduces model computational complexity and improves speed.

The third part, Neck, including the convolutional layer and the C2f module, conducts a multi-scale deep fusion of the features and transmits the features to the predicted output part. The C2f module in the backbone network and Neck part combines the ELAN structure design of YOLOv7 and replaces the C3 module in YOLOv5. Figure 3 depicts the C2f structure with a richer gradient flow, which enhances the feature fusion capability of the CNN and achieves further network structure lightweighting.

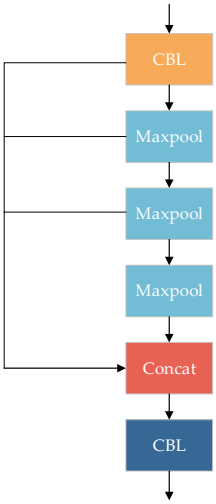


Figure 2. SPPF module.

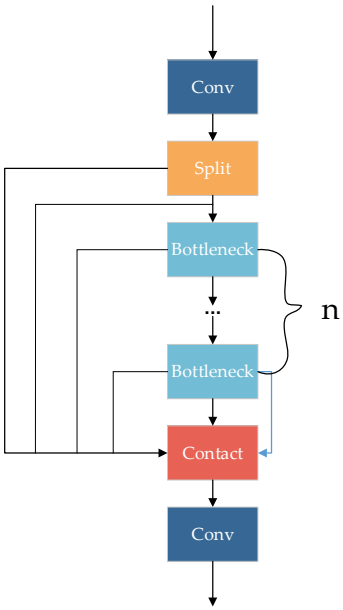


Figure 3. C2f module network structure.

The fourth part, Head, is responsible for predicting output. It predicts the characteristic chart after deep fusion, and outputs the target anchor frame position, target types, and confidence information. Unlike previous YOLO series models, the Head part opts for the decoupling head structure instead of the coupling head structure, and chooses the Anchor-free idea instead of the Anchor-Base idea. Because of the different contents focused by the classification and localization, the decoupling head structure uses different branches for operations, enhancing the detection effect. With these improvements in structure, the YOLOv8 model has seen enhancements in both speed and accuracy.

3. Algorithm Optimization

To solve the problem of the YOLOv8 model in the application of plant pests and diseases recognition, this paper proposes the improved YOLOv8 model to enhance the comprehensive performance in the recognition of plant pests and diseases. The improved YOLOv8 model is shown in Figure 4. In the original YOLOv8 model, this paper introduces improvements in the following areas. Firstly, in the Backbone network, after the SPPF module and before the output feature, the GAM is adopted to weigh the important feature information, improving the accuracy of the model. Secondly, in the mixed feature part

of the Neck, the RFA Conv method is used instead of standard convolution operations to enhance the processing ability of feature information. Thirdly, in terms of datasets, a comprehensive dataset containing various representative plant diseases and pests is constructed to enhance the applicability of the model.

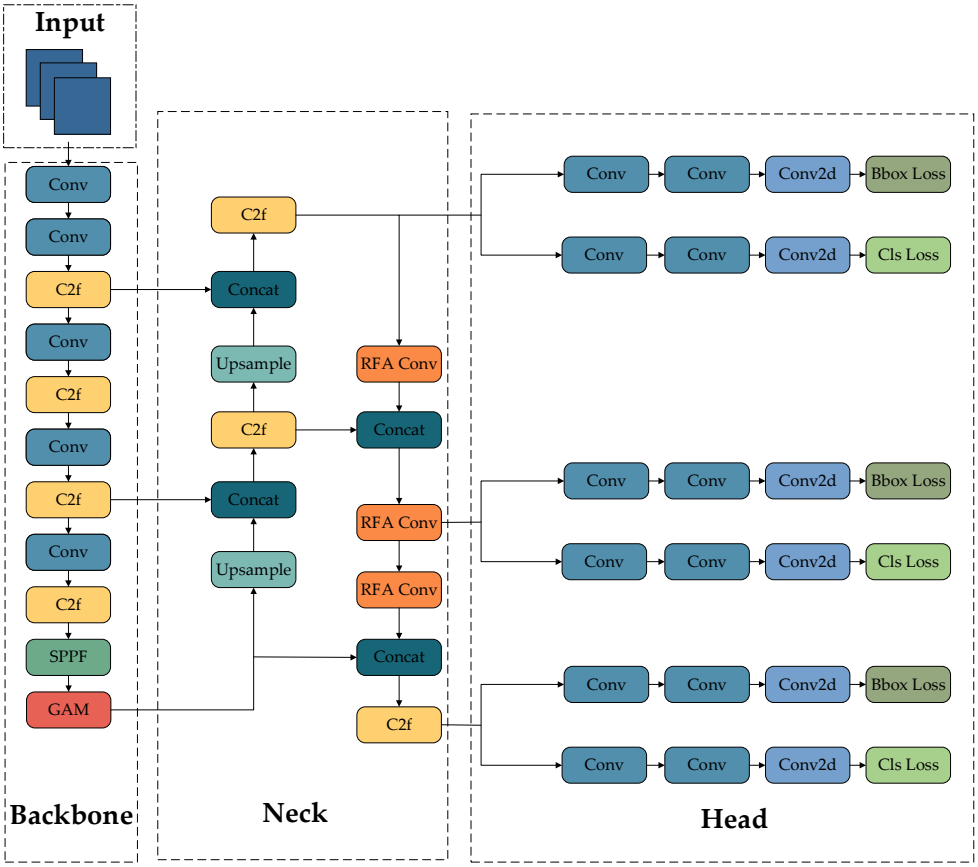


Figure 4. Improved YOLOv8 network structure.

(1) GAM

In tasks such as target detection and image classification, the attention mechanism can improve the ability of the model to locate and recognize important features in images of the model, thereby enhancing detection accuracy. Inspired by human attention, the attention mechanism aims to mimic how the human brain processes key areas in images. Its primary goal is to select target information crucial for the current task from a plethora of information, focus on it, and ignore the background areas that do not match the target features, thereby saving computational resources and improving computational efficiency [25]. At present, the attention mechanism is mainly divided into spatial attention and channel attention. Space attention can improve the accuracy of model positioning and identification in the important areas. Channel attention can enhance the speed and accuracy and the relationship between different channels. In the target detection task, applying spatial attention and channel attention to the outputs of the convolution layer can yield more precise feature information representations. Traditional SE attention mechanisms [26], through learning adaptive channel weights, pay more attention to crucial channel information. Nevertheless, they only consider the attention of the channel dimension, thus they are unable to capture the attention in the spatial dimension, limiting their applicability. The CBAM [27], combining with the convolution and attention mechanism, can pay attention to images from both dimensions of space and channel, but it comes with higher computational complexity.

The GAM improves the performance of deep neural networks by reducing information and amplifying overall interaction representations. The introduction of this optimization

algorithm is based on the proposal of the first-order smoothness theory. Compared to zero-order smoothness, first-order smoothness focuses on the norm of the maximum gradient within the parameter neighborhood, making it more capable of capturing the trend of loss changes. Therefore, the GAM optimizes both prediction errors and the norm of the maximum gradient within the neighborhood during training in Figure 5. Like CBAM, GAM uses space attention and channel attention but handles these dimensions differently. In terms of channel attention, first, it transforms the input feature diagram, passes it through an MLP, and then restores it to the original dimension, and finally performs a Sigmoid function for output.

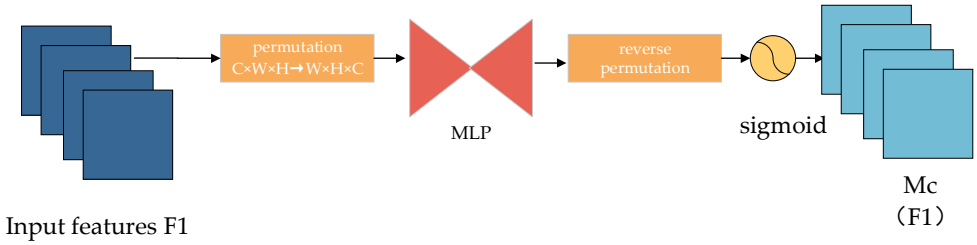


Figure 5. Image of processing by GAM in terms of channel attention.

In terms of spatial attention, the GAM first reduces the number of channels by using a convolutional kernel of size 7 to reduce the calculation amount. Then, it undergoes another convolutional operation with a kernel size of 7 to increase the number of channels, maintaining the consistency of the number of channels. Finally, it gives outputs through a Sigmoid function as shown in Figure 6.

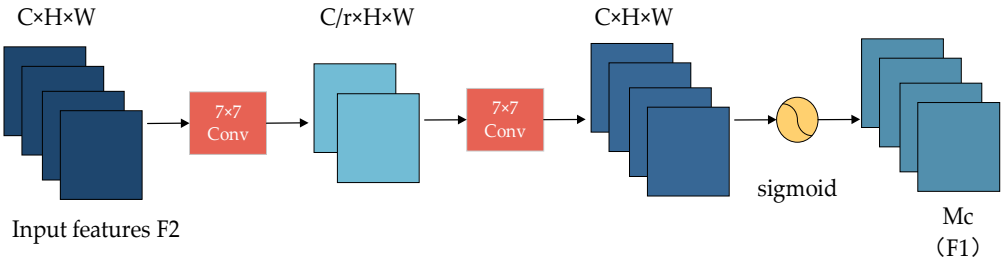


Figure 6. Image of processing by GAM in terms of spatial attention.

(2) RFA Conv module

The existing CBAM and CA attention mechanisms focus only on spatial characteristics in terms of spatial attention and do not completely solve the problem of parameter sharing in convolutional kernels. The RFA attention mechanism not only considers the spatial characteristics of the receptive field but also provides effective attention weights for large-sized convolutional kernels. The receptive field spatial features refer to the feature diagram transformed by the spatial characteristics, which consist of the non-overlapping sliding windows. Each 3×3 size window in the receptive field spatial characteristics represents a receptive field block. Generally speaking, the calculation of RFA can be expressed as follows:

$$F = Softmax\left(g^{i\times i}(AvgPool(X)) \times ReLU\left(Norm\left(g^{k\times k}(X)\right)\right)\right) \quad (1)$$

where $g^{i\times i}$ presents a group convolution with a size of $i \times i$, k represents the size of the convolutional kernels, $Norm$ shows a normalization,X indicates the input feature diagram.

In Figure 7, RFA Conv developed by RFA can replace standard convolutional calculations, effectively reducing computational costs and parameter increments while improving network performance. First of all, it uses the fast Group Conv method to extract the receptive field features, aggregates the global information for each receptive field feature

using AvgPool, and then employs 3 × 3 sets of convolutional operations to interact with information. Finally, it emphasizes the importance of each feature in the receptive field features through softmax.

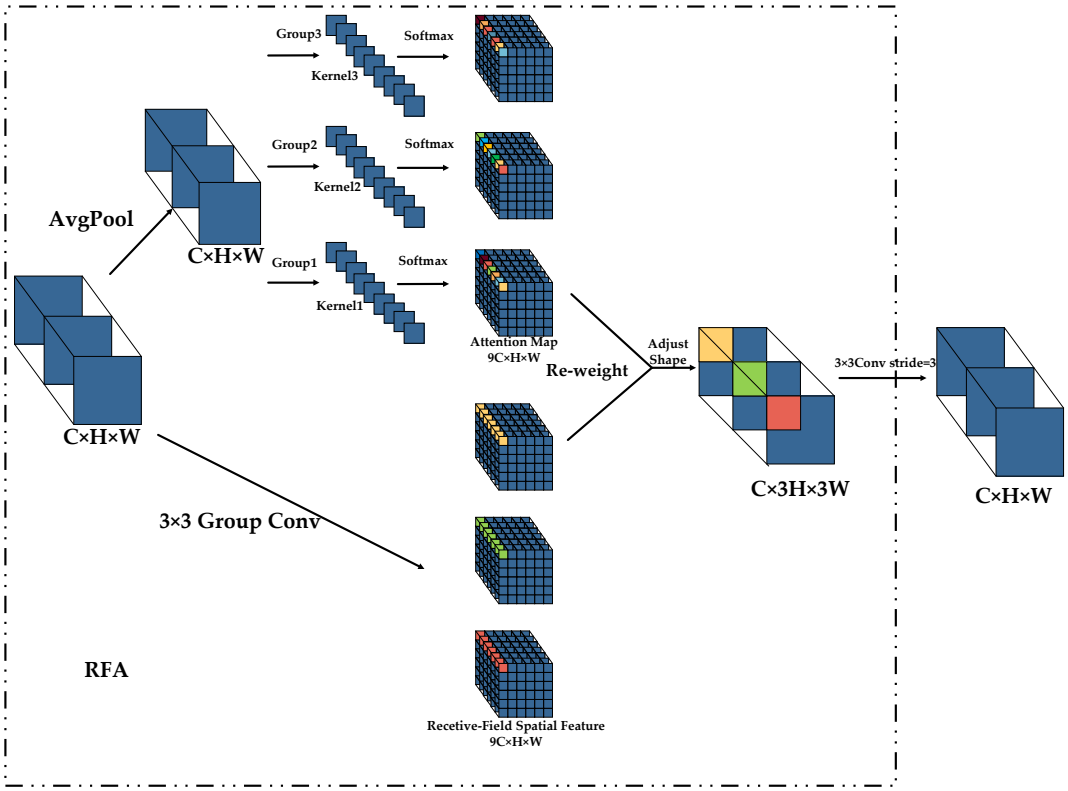


Figure 7. RFA schematic diagram.

4. Experimental Scheme

4.1. Pavement Defect Dataset

Although there is a wide variety of plant species, and their diseases and pests vary, according to the characteristics of plant diseases and pests, the common symptoms can be summarized, and the test recognition can be performed to establish a dataset to improve the model detection accuracy. The pictures in the dataset are collected in real fields, which can enhance the generalizability of the model.

Different plants can experience the same diseases and insect pests. For example, pests such as leaf-eating insects can cause circular transparent spots on leaves. Powdery mildew can lead to tiny white powdery spots on leaves, gradually expanding into dirty white to tan-colored round spots. This dataset includes not only common plant diseases like brown spot, and powdery mildew, but also specific diseases unique to certain plant varieties, such as cotton leaf curl disease.

During the experiment, for the purpose of testing the accuracy, speed, generalization, and other performance aspects of the model, the dataset in this paper selected rice as a representative of grain crops, and cotton as a representative of economic crops. The pictures are collected and arranged from the Kaggle platform (https://www.kaggle.com/, 1 May 2024), which is an online community for data scientists. The specific categories and quantities are shown in Table 1 below. In the experimental process, the rice dataset is divided into training sets, verification sets, and test sets in a ratio of 8:1:1, while the ratio of the cotton dataset is 6:2:2.

Table 1. Specific categories of diseases and pests and quantities.

Plant Categories	Categories	Codes	Quantities	Image Examples
rice	Blast	S0	18,239	
	Blight	S1	20,498	
	Brown spot	S2	32,172	
cotton	Alternaria Leaf Spot	M0	92	
	Curl Leaves	M1	90	
	Foliar disease	M2	541	
	Herbicide	M3	761	

4.2. Experimental Process

With the intention of verifying the performance of the improved model in various aspects, this paper sets up a comparative experiment to show the superiority of the improved model and sets out the ablation experiments to prove the feasibility of the improvement strategy. Aimed to ensure the rigor and accuracy of the experiment, similar environmental parameters and model parameters are set up as much as possible. The development language of this model is mainly Python 3.9. The operating system is Linux. The graphics card is Tesla V100S PCIe 32 GB (NVIDIA, Santa Clara, CA, USA). The Batch Size is set to 16. The training loss variation curves are shown in Figure 8 below, where the model gradually converges with increasing iterations and reaches stability after 150 rounds of training. The training results are used as the final weight parameters.

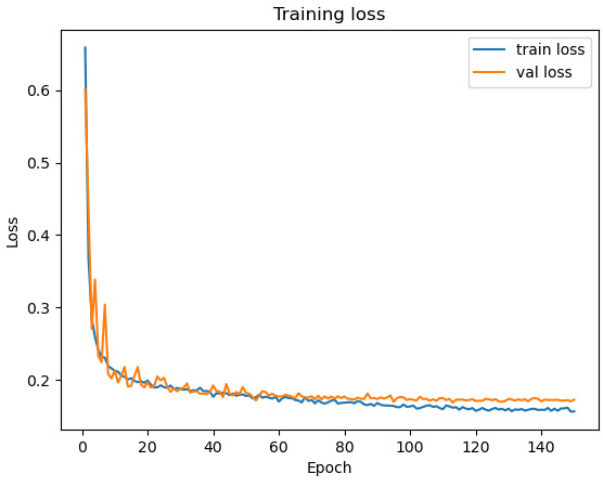


Figure 8. Training loss variation curves.

During the ablation experiments, this study adopts the YOLOv8 model with no improvement, partial improvement, and complete improvement for recognition and detection, and finally employs comprehensive comparative analysis of the detection results of all situations.

In the comparative experimental stage, this study uses the Faster R-CNN model in the One-stage detection algorithm, the YOLOv7 and YOLOv8 models in the Two-stage target detection algorithm, and the improved YOLOv8 model for comparison experiments.

4.3. Experimental Evaluation Scheme

In order to comprehensively evaluate the performance of the model, this paper chooses commonly used evaluation metrics of the YOLO algorithm, including precision, recall rate, average precision (AP), mean average precision, and F1 score. These metrics are based on many predicted results, which are statistically classified and calculated. The predicted results in the statistical process can be divided into four categories: true positive (*TP*) for correctly predicting positive samples, false negative (*FN*) for wrongly predicting negative samples, true negative (*TN*) for correctly predicting negative samples, and false positive (*FP*) for wrongly predicting positive samples. Here, positive samples represent targets, while negative samples represent backgrounds.

(1) Precision

The precision means the proportion of true positives among all samples classified as positive, which indicates the accuracy of the predictions. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{2}$$

(2) Recall rate

The recall rate indicates the proportion of being correctly classified as true positives among all actual positive samples, indicating whether detected targets are complete or not. The calculation formula is as follows:

$$R = \frac{TP}{TP + FN} \tag{3}$$

(3) AP

The average precision can evaluate the comprehensive performance of the detection accuracy and recall rate of the model on different categories. Based on the area under the precision-recall (*PR*) curve, it is calculated. The value of the average accuracy ranges from 0 to 1, where a higher value indicates a better detection performance.

(4) mAP

As an important indicator for measuring the performance of the target detection algorithm, the mean average precision indicates the comprehensive calculation of precision at different recall rates for each category. The calculation formula is as follows: n represents the number of categories, m represents the number of targets under this category, and $P(r)$ means that the precision value of the recall rate is R. The higher the mAP value, the better the performance of the algorithm in detecting targets of each category.

$$mAP = \frac{1}{n}\sum\left[\frac{1}{m}\sum P(r)\right] \tag{4}$$

(5) F1 score

The F1 score considers both precision and recall, and it is the harmonic mean of precision and recall. The calculation formula is as follows:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{5}$$

The above indicators can evaluate the model in terms of detection accuracy, positioning accuracy, and other aspects, thereby expressing the overall performance of the model. In addition, the FPS index is required to evaluate the detection speed of the model. The FPS value represents the number of images that the model can process per second.

5. Analysis of Experimental Results

5.1. Ablation Experiment

To verify the application effect of the improvement strategies mentioned earlier, this paper sets four groups of experiments based on the principle of ablation to analyze the role of different modules on model improvement. This experiment is conducted on two datasets of rice and cotton, and all experimental groups have the same environmental parameters and model parameters. The experimental results are shown in Table 2 below, indicating that the improved model has different enhancement effects in both detection accuracy and detection speed.

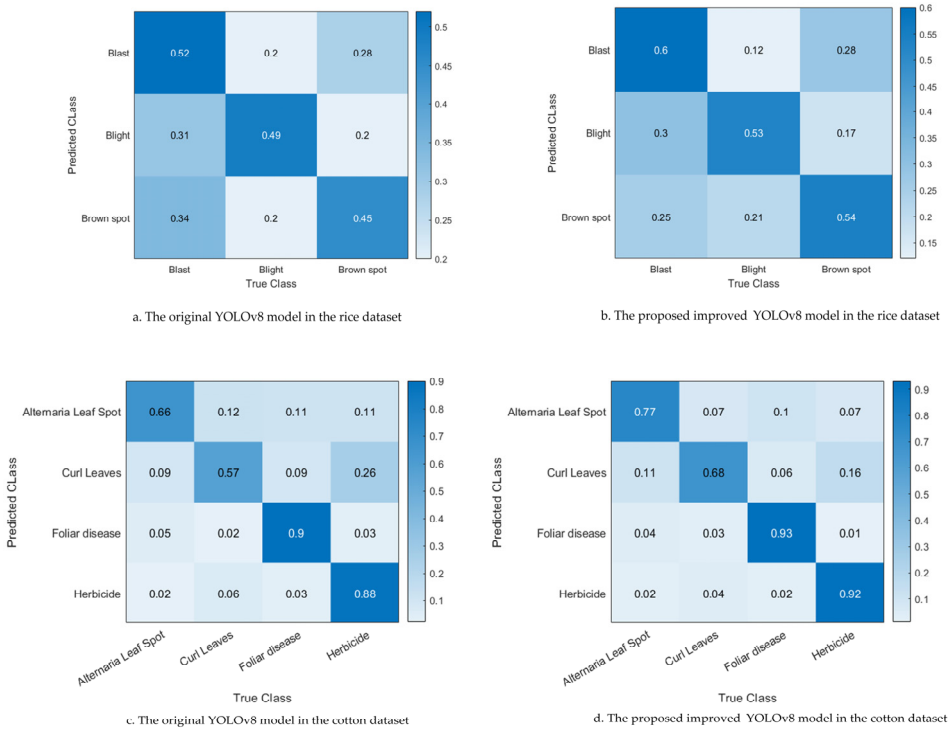
Table 2. Results of the ablation experiment. The "×" means that the module is not added. The "√" means that the module is added in the model.

Model	GAM	RFA Conv	Rice			Cotton		
			mAP	F1	FPS	mAP	F1	FPS
YOLOv8	×	×	66.47%	0.47	57.18	79.56%	0.71	57.82
Version 1	√	×	67.66%	0.51	56.45	80.91%	0.75	56.76
Version 2	×	√	68.04%	0.48	56.10	80.01%	0.72	56.32
Version 3	√	√	71.27%	0.53	55.57	82.91%	0.77	55.65

The ablation experiments can demonstrate the performance of the proposed model in this paper under different improvement states. In Version 1, the GAM module is enabled, and inserted after the SPPF module in the Backbone network. Its main role is to weigh important feature information, thereby improving detection accuracy. The experimental results demonstrate that the model in Version 1 improves the detection accuracy on both datasets, with better performance on the cotton dataset with complex disease and pest categories. However, there is a loss in detection speed. Version 2 enables the RFA Conv module, that is, in the mixed feature part of the Neck, the RFA Conv method replaces standard convolution operations. Its main function lies in the reduction of calculation costs, thereby increasing the calculation speed. The experimental results show that the model in Version 2 improves the calculation speed on both datasets. Version 3, building on Version 2, enables the GAM to improve the accuracy of calculation while maintaining

the calculation speed. According to experimental data, the above-mentioned improved modules have different degrees of improvement in the accuracy and speed of the model. The improved model mAP values reach 65.97% in the rice dataset and 82.91% in the cotton dataset, representing an average improvement of 3.3% compared to the original model. The average F1 score reaches 0.53 in the rice dataset and reaches 0.77 in the cotton dataset, having an average increase of 0.055. The improvements in the above indicators show that the proposed model in this paper has strong reference significance for improving the accuracy of recognition models of plant pests and diseases.

To assess the reliability of the test results, this paper set confusion matrices to analyze the proposed improved YOLOv8 model and the original YOLOv8 model. The results of the confusion matrices are shown in Figure 9.



a. The original YOLOv8 model in the rice dataset

b. The proposed improved YOLOv8 model in the rice dataset

c. The original YOLOv8 model in the cotton dataset

d. The proposed improved YOLOv8 model in the cotton dataset

Figure 9. The confusion matrices generated by the proposed improved YOLOv8 model and original YOLOv8 model.

The confusion matrix consists of row vectors representing the predicted class and column vectors representing the true class. Upon examining the confusion matrix of the proposed improved YOLOv8 model, it is evident that the recall rate experienced a significant increase compared to that of the original YOLOv8 model. Specifically, in the rice dataset, the average recall rate saw an improvement of 8.51%. However, there are a few classification errors in the blast and brown spot categories, possibly due to the misjudgments due to the similarity in shape and color of these two diseases. In the cotton dataset, the majority of categories were correctly detected, with the recall rates reaching 0.93 and 0.92 in the foliar disease and herbicide categories, respectively. Overall, the improved model demonstrates a relatively reliable ability to extract features from images, resulting in improved detection accuracy.

5.2. Comparative Experiment

With the aim of verifying the significant advantages of the improved model proposed in this paper compared to other models, multiple groups of comparative experiments are set up. The Faster R-CNN, YOLOv7, YOLOv8, and the improved YOLOv8 are employed to rice and cotton datasets, and a comparative analysis of visualization results and evaluation

results is performed. To ensure the reliability of the comparison results, all experimental parameters are kept the same, along with identical experimental environment settings.

In the comparison stage of the visualization results, each typical disease and pest is tested by experimental models. Different models present different effects on the detection of typical diseases and insect pests. The results are shown in Figure 10.

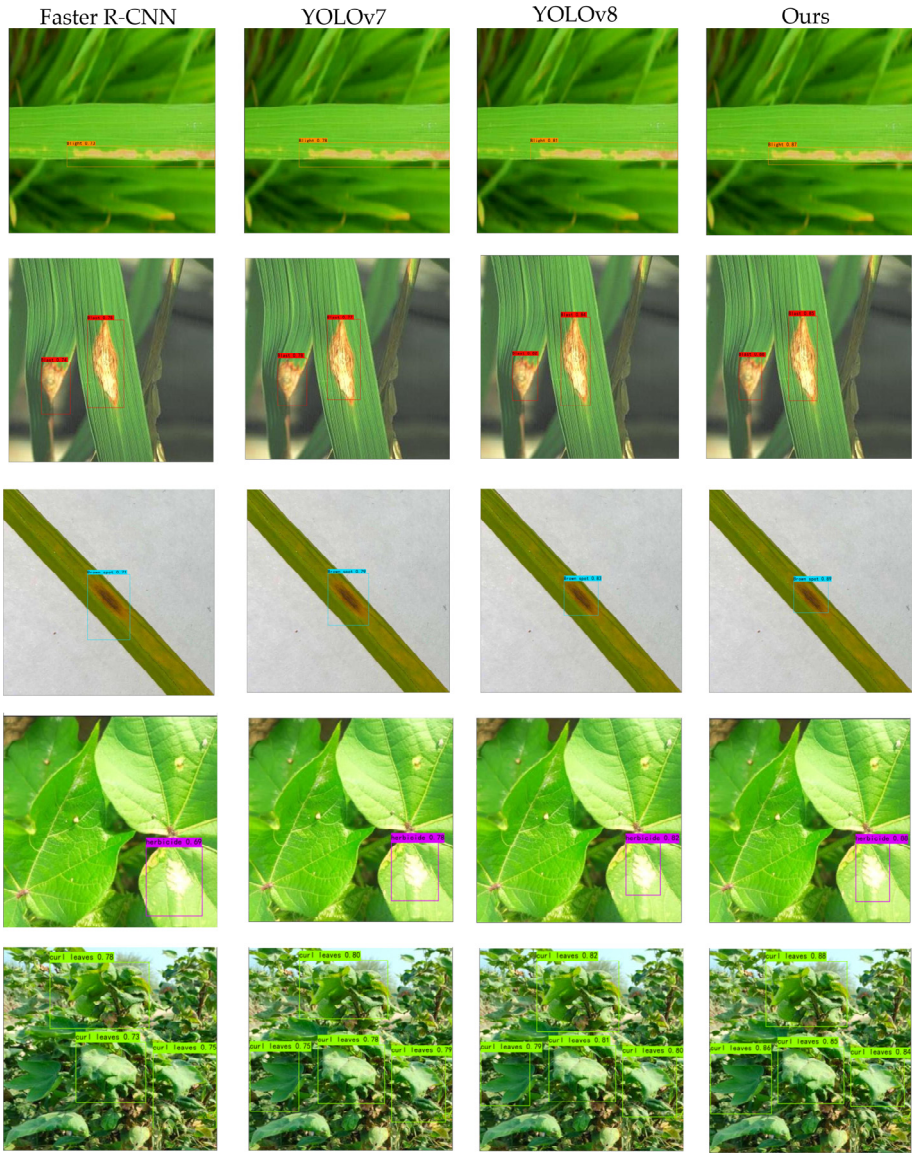


Figure 10. Recognition results of different models for typical diseases and pests.

Figure 9 indicates the testing results of each model in different typical diseases and pest recognition tasks, respectively. Through the comparison analysis of the experiment, the overall performance of the Faster R-CNN model is poor. The main problems include the low matching accuracy of predicted bounding boxes and instances of missed detections. The model also exhibits different detection effects on the rice and cotton datasets. Compared to the Faster R-CNN, YOLOv7 has a better overall performance, but there are still some missed examinations, and the performance is unstable during the testing of different disease and pest categories. YOLOv8 shows good performance, but it has lower confidence levels for certain disease and pest categories, and its results for detecting small targets are subpar. Based on analysis of typical and numerous samples, the improved YOLOv8 proposed in this paper shows higher confidence in target detection, better matching accuracy of predicted boxes, fewer missed detections, and more comprehensive detection of small

targets compared to other detection models. In conclusion, the improved model proposed in this paper demonstrates superior visual detection results.

In the comparison stage of the evaluation results, mAP values, F1 values, and FPS values are used to measure the performance of different models. The results are shown in Table 3.

Table 3. Test results of different models.

Model	Rice			Cotton		
	mAP	F1	FPS	mAP	F1	FPS
Faster R-CNN	49.33%	0.45	22.78	76.88%	0.65	23.58
YOLOv7	61.80%	0.49	57.20	78.36%	0.66	57.89
YOLOv8	66.47%	0.47	57.18	79.56%	0.71	57.82
Ours	71.27%	0.53	55.57	82.91%	0.78	55.65

Analyzing the above, the Faster R-CNN performs poorly compared to other models in terms of accuracy and speed. Compared to YOLOv8, YOLOv7 has faster detection speeds but lower detection accuracy, and its overall performance is not outstanding. This is because the YOLOv8 model has a more complex network structure with more calculated parameters, resulting in a loss of detection speed while maintaining detection accuracy. From the perspective of the evaluation indicators and comparing the improvement percentages, in the rice dataset, the improved model in this paper achieves a 7.13% increase in mAP compared to YOLOv8 and a 44.48% increase compared to the Faster R-CNN. The F1 score is improved by 13.78% over YOLOv8 and 17.77% over the Faster R-CNN. In the cotton dataset, the improved model in this paper achieves a 4.21% increase in mAP compared to YOLOv8 and a 7.84% increase compared to Faster R-CNN. The F1 score is improved by 9.86% over YOLOv8 and 20% over the Faster R-CNN. The above computed result fully demonstrates that the proposed model in this paper makes significant advancements in detection accuracy compared to other traditional mainstream methods.

In summary, combining with visualization results and evaluation results, the proposed model in this paper outperforms traditional YOLO models notably in small defect detection and accurate matching of detection boxes. It exhibits higher detection accuracy and efficiency, demonstrating certain superiority.

6. Conclusions

For the sake of achieving accurate and efficient testing and identifying the mission of plant diseases and pests, this paper proposes an improvement method based on the YOLOv8 model. The superiority of the proposed method is verified by the analysis of experimental data. The main research contents of this paper are as follows.

(1) In the Backbone network, after the SPPF module and before the output features, the GAM is adopted to weigh the important feature information, thereby improving the accuracy of the model.
(2) In the mixed feature part of the Neck, the RFA Conv method is used instead of standard convolution operations to enhance the processing ability of feature information, reduce computational complexity and costs, and improve network performance.
(3) In terms of data calculation, the improved model is concentrated in rice datasets and cotton data. The mAP values reach 71.27% and 82.91%. In the ablation and comparison experiments, compared to other models, it shows lower missed detection rates, higher target matching rates, and higher detection accuracy.

Therefore, the application of the improved model proposed in this paper to plant pests and diseases is of great significance for promoting high-quality agricultural development.

Author Contributions: Methodology, Y.W.; Validation, C.Y.; Formal analysis, T.H.; Data curation, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (52205537), Hubei Province Key Research and Development Plan (2021BAA194), and Guangxi Key Research and Development Plan (Guike AB21075009).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: These data were derived from the following resources available in the public domain: https://www.kaggle.com/ (accessed on 1 May 2024).

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

References

1. Meena, S.D.; Susank, M.; Guttula, T.; Chandana, S.H.; Sheela, J. Crop Yield Improvement with Weeds, Pest and Disease Detection. *Procedia Comput. Sci.* **2023**, *218*, 2369–2382. [CrossRef]
2. Food and Agricultural Organization (FAO). Crop Production and Natural Resource Use n.d. Available online: http://www.fao.org/3/y4252e/y4252e06.htm (accessed on 1 May 2024).
3. Chen, P.; Xiao, Q.; Zhang, J.; Xie, C.; Wang, B. Occurrence prediction of cotton pests and diseases by bidirectional long short-term memory networks with climate and atmosphere circulation. *Comput. Electron. Agric.* **2020**, *176*, 105612. [CrossRef]
4. Singh, B.K.; Delgado-Baquerizo, M.; Egidi, E.; Guirado, E.; Leach, J.E.; Liu, H.; Trivedi, P. Climate change impacts on plant pathogens, food security and paths forward. *Nat. Rev. Microbiol.* **2023**, *21*, 640–656. [CrossRef]
5. Jia, S.; Gao, H. Review of Crop Disease and Pest Image Recognition Technology. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *799*, 012045. [CrossRef]
6. Ngugi, L.C.; Abelwahab, M.; Abo-Zahhad, M. Recent advances in image processing techniques for automated leaf pest and disease recognition—A review. *Inf. Process. Agric.* **2020**, *8*, 27–51. [CrossRef]
7. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 142–158. [CrossRef] [PubMed]
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
9. 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; In *Proceedings, Part I 14*. Springer International Publishing: Berlin/Heidelberg, Germany, 2016.
10. Jiao, L.; Dong, S.; Zhang, S.; Xie, C.; Wang, H. AF-RCNN: An anchor-free convolutional neural network for multi-categories agricultural pest detection. *Comput. Electron. Agric.* **2020**, *174*, 105522. [CrossRef]
11. Cynthia, S.T.; Hossain KM, S.; Hasan, M.N.; Asaduzzaman, M.; Das, A.K. Automated Detection of Plant Diseases Using Image Processing and Faster R-CNN Algorithm. In Proceedings of the 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 24–25 December 2019.
12. Tian, Y.; Wang, S.; Li, E.; Yang, G.; Liang, Z.; Tan, M. MD-YOLO: Multi-scale Dense YOLO for small target pest detection. *Comput. Electron. Agric.* **2023**, *213*, 108233. [CrossRef]
13. Yu, J.; Bai, Y.; Yang, S.; Ning, J. Stolon-YOLO: A detecting method for stolon of strawberry seedling in glass greenhouse. *Computers and Electronics in Agriculture* **2023**, *215*, 108447. [CrossRef]
14. Liu, J.; Wang, X. Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. *Front Plant Sci.* **2020**, *11*, 898. [CrossRef] [PubMed]
15. Kumar, V.S.; Jaganathan, M.; Viswanathan, A.; Umamaheswari, M.; Vignesh, J.J.E.R.C. Rice leaf disease detection based on bidirectional feature attention pyramid network with YOLO v5 model. *Environ. Res. Commun.* **2023**, *5*, 065014. [CrossRef]
16. Luo, D.; Xue, Y.; Deng, X.; Yang, B.; Chen, H.; Mo, Z. Citrus Diseases and Pests Detection Model Based on Self-Attention YOLOV8. *IEEE Access* **2023**, *11*, 139872–139881. [CrossRef]
17. Yang, Y.; Di, J.; Liu, G.; Wang, J. Rice Pest Recognition Method Based on Improved YOLOv8. In Proceedings of the 2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 12–14 January 2024; pp. 418–422.
18. Di, Y.; Phung, S.L.; Van Den Berg, J.; Clissold, J.; Bouzerdoum, A. TP-YOLO: A Lightweight Attention-Based Architecture for Tiny Pest Detection. In Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023; pp. 3394–3398.
19. Uddin, M.S.; Mazumder MK, A.; Prity, A.J.; Mridha, M.F.; Alfarhood, S.; Safran, M.; Che, D. Cauli-Det: Enhancing cauliflower disease detection with modified YOLOv8. *Front. Plant Sci.* **2024**, *15*, 1373590. [CrossRef] [PubMed]
20. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLO (Version 8.0.0) [Computer Software]. 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 1 May 2024).

21. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. Computer Vision and Pattern Recognition. *arXiv* **2021**, arXiv:2112.05561.
22. Zhang, X.; Liu, C.; Yang, D.; Song, T.; Ye, Y.; Li, K.; Song, Y. RFA Conv: Innovating Spatial Attention and Standard Convolutional Operation. Computer Vision and Pattern Recognition. *arXiv* **2023**, arXiv:2304.03198.
23. Iqbal, S. *Insect, Pest and Disease Management in Rice*; Austin Publishing Group: Irving, TX, USA, 2020.
24. Chohan, S.; Perveen, R.; Abid, M.; Tahir, M.N.; Sajid, M. Cotton Diseases and Their Management. In *Cotton Production and Uses*; Ahmad, S., Hasanuzzaman, M., Eds.; Springer: Singapore, 2020.
25. Obeso, A.M.; Benois-Pineau, J.; Vázquez, M.S.G.; Acosta, A.R. Visual vs internal attention mechanisms in deep neural networks for image classification and object detection. *Pattern Recognit.* **2022**, *123*, 108411. [CrossRef]
26. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

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.