Article

YOLOv11-RCDWD: A New Efficient Model for Detecting Maize Leaf Diseases Based on the Improved YOLOv11

Jie He, Yi Ren, Weibin Li and Wenlin Fu

Topic

Article

YOLOv11-RCDWD: A New Efficient Model for Detecting Maize Leaf Diseases Based on the Improved YOLOv11

Jie He [1,2] , Yi Ren [1,2] , Weibin Li [1,2,*] and Wenlin Fu [1,2]

1   Laboratory of AI, Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China; hejie@xidian.edu.cn (J.H.); yiren4323@gmail.com (Y.R.); fuwenlin@xidian.edu.cn (W.F.)
2   School of Artificial Intelligence, Xidian University, Xi'an 710071, China
*   Correspondence: weibinli@xidian.edu.cn

Abstract: Detecting pests and diseases on maize leaves is challenging. This is especially true under complex conditions, such as variable lighting and occlusion. Current methods suffer from low detection accuracy. They also lack sufficient real-time performance. Hence, this study introduces the lightweight detection method YOLOv11-RCDWD based on an improved YOLOv11 model. The proposed approach enhances the YOLOv11 model by incorporating the RepLKNet module as the backbone, which significantly enhances the model's capacity to capture characteristics of maize leaf pests and diseases. Additionally, the CBAM is embedded within the neck feature extraction network to further refine the feature representation to augment the model's capability to identify and select essential features by introducing attention mechanisms in both the channel and spatial dimensions, thereby improving the accuracy of feature expression. We have also improved the model by incorporating the DynamicHead module, WIoU loss function, and DynamicATSS label assignment strategy, which collectively enhance detection accuracy, efficiency, and robustness through optimized attention mechanisms, better handling of low-quality samples, and dynamic sample selection during training. The experimental findings indicate that the improved YOLOv11-RCDWD model effectively detected pests and diseases on maize leaves. The precision reached 92.6%, while the recall was 85.4%. The F1 score was 88.9%, and the mAP@0.5 and mAP@0.5~0.95 demonstrated an improvement of 4.9% and 9.0% over the baseline YOLOv11s. Notably, the YOLOv11-RCDWD model significantly outperformed other architectures such as Faster R-CNN, SSD, and various models within the YOLO series, demonstrating superior capabilities in terms of detection speed, parameter count, computational efficiency, and memory utilization. This model achieves an optimal balance between detection performance and resource efficiency. Overall, the improved YOLOv11-RCDWD model significantly reduces detection time and memory usage while maintaining high detection accuracy, supporting the automated detection of maize pests and diseases, and offering a robust solution for intelligent monitoring of agricultural pests.

Keywords: maize leaf pest and disease detection; YOLOv11s-RCDWD; RepLKNet; CBAM; object detection

1. Introduction

Maize (*Zea mays* L.) is widely regarded as one of the most significant food crops globally, with its yield and quality directly affecting food security and economic development [1,2]. However, maize is vulnerable to various diseases and pests throughout its growth, such as maize rust and leaf spot, severely affecting its yield and quality [3]. Therefore, accurate disease identification is critical for disease control and yield assurance

in maize cultivation [4]. Since disease symptoms predominantly appear on the leaves, leaf images become a vital basis for disease identification [5]. Traditional methods to detect pests and diseases in maize leaves rely on manual observation. These methods are inefficient, burdensome, and subjective, and they cannot meet the demands of modern agriculture for refined and large-scale production [6]. Therefore, there is a critical need to explore efficient and accurate methodologies for the detection of maize pests and diseases. This endeavor is vital to address the limitations associated with traditional techniques and to facilitate rapid and precise assessments of pest infestations and disease occurrence [7].

With the recent rapid advancements in computer vision technology, machine learning and deep learning techniques have found extensive application across various domains. They are particularly effective in agricultural monitoring, especially for the detection of pests and diseases [8,9]. For example, Panigrahi et al. [10] compared the accuracy of five standard machine learning techniques. The methodologies examined included NB, DT, KNN, SVM, and the frequently implemented RF for the detection of pests and diseases affecting maize crops. They found that the RF method achieved a high accuracy of 79.23%. Paul et al. [11] developed a mobile application based on the pre-trained VGG16 architecture using a convolutional neural network (CNN), facilitating the efficient identification and classification of maize leaf diseases.

The YOLO series has emerged as a pivotal method among numerous detection algorithms, due to its outstanding real-time performance and high level of detection accuracy [12,13]. From YOLOv1 to the latest YOLOv11, the YOLO family has been continuously optimized for detection accuracy, speed, and model complexity, providing robust technical support for agricultural detection, especially intelligent detection of pests and disease [14–17]. For instance, Fan et al. [18] combined YOLOv5 with dark channel enhancement for strawberry ripeness recognition and achieved a test accuracy exceeding 90%. Wang et al. [19] introduced a detection method for wheat seedlings, which improved YOLOv5 by replacing global annotation with local annotation, adding a micro-scale detection layer and a spatial depth convolution module. These modifications significantly enhanced the extraction capability of small features and increased detection accuracy to 90.1%. Li et al. [20] proposed YOLO-Leaf for apple leaf disease detection, utilizing DSConv to achieve robust feature extraction, BiFormer for enhanced attention mechanisms, and IF-CIoU to optimize bounding box regression. YOLO-Leaf outperformed existing models in terms of detection accuracy. Lu et al. [21] developed the cotton boll detection model COTTON-YOLO that improved YOLOv8n by introducing the C2F-CBAM module and the Gold-YOLO neck architecture designed for enhanced information flow and enhanced feature integration. Yang et al. [22] identified maize leaf spots by introducing the Slim-neck and GAM attention in YOLOv8. This setup significantly improved the model's recognition ability for maize leaf spots, demonstrating a 3.79% improvement in accuracy (P) and 4.65% in recall (R) compared with the original YOLOv8 model.

Despite advancements in detecting pests and diseases in maize leaves, existing methods still struggle with complex agricultural scenarios. For instance, occlusion and overlapping of maize leaves lead to missed and false detections [23], while variable lighting conditions can significantly degrade detection performance [24]. Moreover, the classification accuracy of existing methods is inadequate for differentiating between pests and diseases with similar morphologies [25]. The YOLO series algorithms, widely used for object detection, also exhibit limitations in this context. Their complex multi-branch backbone structure hampers the speed of disease detection and leads to underperformance when processing small and dense targets in complex environments [26]. Additionally, these methods rely heavily on predefined anchor boxes, resulting in slower inference speeds. The rigidity of sample allocation strategies further reduces the efficient utilization

of features during training. To address these challenges, we propose a lightweight and efficient solution tailored for mobile agricultural environments. Our model aims to improve detection accuracy, enhance robustness under varying conditions, and optimize computational efficiency. This not only holds practical significance for real-time monitoring and management in agriculture but also contributes to the scientific development of efficient detection algorithms for use in resource-constrained settings.

Motivated by the concerns outlined above, this study proposes a high-performance and lightweight model for detecting maize leaf diseases, named YOLOv11s-RCDWD (YOLOv11s-RepLKNet-CBAM-DynamicHead-WIoU-DynamicATSS). The proposed model is built upon the latest YOLO series object detection algorithm, YOLOv11s, with the aim of accurately and efficiently locating and detecting maize leaf diseases in natural scenes with complex backgrounds. Furthermore, this model is suitable for deploying on mobile devices.

The main contributions of this study are as follows:

1. In the backbone layer, we replace the C3k2 backbone network with the RepLKNet network, improving detection performance and efficiency through re-parameterization;
2. Introducing the CBAM (Selective Kernel Attention, CBAM) attention module to enhance the model's ability to capture multi-scale features. The CBAM allows the model to focus on key features of diseased maize leaves, thereby improving feature extraction capability;
3. Adopting the DynamicHead detection head to unify the scale-aware, space-aware, and task-aware aspects of object detection by integrating multiple attention mechanisms. This strategy significantly improves the representation ability of the detection head without increasing computational overhead;
4. Optimizing the loss function using the symmetric IoU loss function WIoU to improve the bounding box regression accuracy and enhance the model's localization capability;
5. Replacing the original fixed-ratio positive and negative sample allocation strategy with DynamicATSS to adjust the selection mechanism for positive and negative samples based on statistical information obtained during training. DynamicATSS improves the model's generalization ability.

2. Materials and Methods

2.1. Production of Datasets

This study constructed a sample dataset covering multiple scenarios and various typical corn diseases, based on an open-source dataset (released by Plant Village [27]) and a self-built dataset (from local corn planting bases in Yangling, Shaanxi), to comprehensively evaluate the performance of YOLOv11s-RCDWD in detecting corn leaf diseases.

LabelImg 1.8.1 annotation software was used to annotate the collected sample images accurately [28]. Specifically, the disease-affected areas in the images were completely enclosed using the maximum horizontal rectangular bounding boxes, and the annotations were stored in VOC-format XML files. Each annotated image underwent data augmentation; first, the image resolution in the training set was uniformly adjusted to 640×640 pixels to generate clear and standardized image samples. Then, randomly flipping the images vertically and horizontally increased the samples' diversity. Finally, the images were normalized based on their mean and standard deviation to eliminate differences in lighting and contrast. The VOC-format XML annotation files were then converted into YOLO-format TXT files, serving as input data for model training. These operations effectively simulated the growth posture, lighting conditions, and shooting angle variations of corn in real-world scenarios, enhancing the diversity of the training samples. A total of 7928 labeled corn disease images were successfully constructed, including 1510 images of rust, 1275 images of gray leaf spot, 1305 images of northern leaf blight, 1163 images of southern leaf blight

images, 1260 images of round spot disease, and 1415 images of healthy corn leaves (Figure 1). These images exhibited significant differences in morphology, texture, and color, providing rich feature information for disease identification. The augmented images were randomly divided into training, testing, and validation sets based on a 7:2:1 ratio (5549, 1586, and 793, respectively) for subsequent model training and testing.
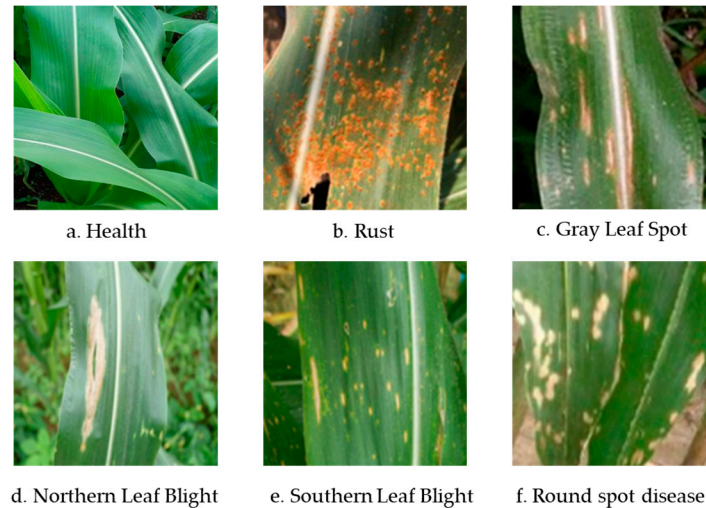

a. Health	b. Rust	c. Gray Leaf Spot
d. Northern Leaf Blight	e. Southern Leaf Blight	f. Round spot disease

Figure 1. Original dataset: (**a**) healthy, (**b**) rust, (**c**) gray leaf spot, (**d**) northern leaf blight, (**e**) southern leaf blight, and (**f**) round spot disease.

2.2. Model Improvement

2.2.1. Improved YOLOv11 Network Model Construction

YOLOv11 is the latest YOLO version, inheriting and refining the core concepts and technical frameworks of its predecessors to enhance detection speed, accuracy, and robustness [29]. Significant improvements have been made in its backbone network and neck layers. Addressing the current challenges of large model parameters, large file size, and low detection accuracy in maize leaf disease detection in complex scenarios, this study enhances the YOLOv11s model.

Specifically, the backbone network uses RepLKNet as the feature extraction network. RepLKNet utilizes advanced convolution operations and re-parameterization techniques to significantly reduce computational load while maintaining feature extraction capabilities, thereby enhancing the model's lightweight characteristics. In addition to its foundational architecture, RepLKNet incorporates the Partial Spatial Attention (PSA) module, which directs the model's focus toward critical feature regions. Furthermore, the neck network integrates the CBAM, an attention mechanism that enhances the model's ability to capture multi-scale features by combining both channel and spatial attention strategies. This dual approach enables a concentrated emphasis on key features related to pests and diseases. Moreover, the neck network aggregates features of different scales through the ELAN-W module, which employs the SPPCSPC module for spatial pyramid pooling, further improving the handling of multi-scale features. The detection head incorporates the DynamicHead module, which integrates various attention mechanisms to unify scale, spatial, and task awareness during object detection, significantly boosting the representational capacity of the detection head without additional computational overhead. The proposed Adaptive Spatial Feature Fusion (ASFF) module adaptively fuses shallow and deep features to reduce variance in the feature scale, accurately detecting large, medium, and small targets. The accuracy of bounding box regression is improved by optimizing the loss function through the Symmetric IoU Loss Function (WIoU). WIoU optimizes the precision of bounding box localization, thereby enhancing the model's overall detection performance.

Lastly, the dynamic adjustment mechanism (DynamicATSS) dynamically selects positive and negative samples based on statistical information obtained during training, replacing the traditional fixed-ratio allocation strategy and significantly improving the model's generalization capabilities.

Figure 2 presents improved YOLOv11s-RCDWD model. Through the optimizations presented above, our model excels in maize leaf disease and pest detection, achieving high precision and performance in locating and identifying diseases and pests while maintaining excellent lightweight characteristics suitable for deployment on agricultural mobile devices.
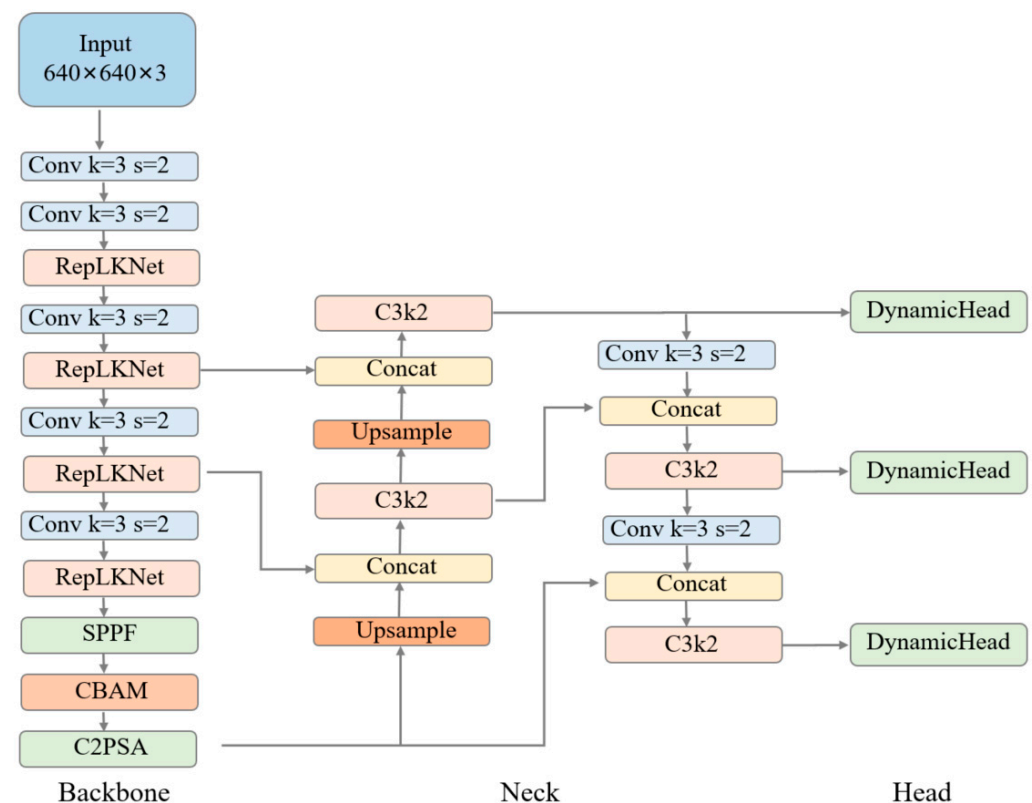

Figure 2. Improved YOLOv11 model structure YOLOv11s-RCDWD.

2.2.2. RepLKNet

In complex scenarios characterized by interference and overlapping foliage, the textural features of diseased areas on maize leaves exhibit a high degree of complexity, posing significant challenges for accurate detection of these diseased regions. Therefore, this study introduces RepLKNet as the backbone network to enhance the model's precision in detecting maize diseases in contexts of overlapping and occluded backgrounds [30]. RepLKNet significantly improves feature extraction performance by expanding the effective receptive field (ERF) and enhancing the model's capability to extract shape information. The ERF, based on a quantitative metric proposed by Luo et al. [31], characterizes the contribution of each input pixel within the receptive field to the output of the n-th layer unit in the network. Figure 3 depicts RepLKNet, which primarily comprises three core modules: Stem, Stage, and Transition. The first of these is responsible for initial feature extraction, the second achieves multi-level feature learning by cascading multiple large kernel convolution sub-modules, and the third down-samples feature maps and adjusts the number of channels. Using RepLKNet with a larger effective receptive field allows the model to effectively extract global contextual information and local detail features from input images, demonstrating significant advantages, particularly when used with complex scenes of maize leaf diseases including shadow interference and overlapping foliage.
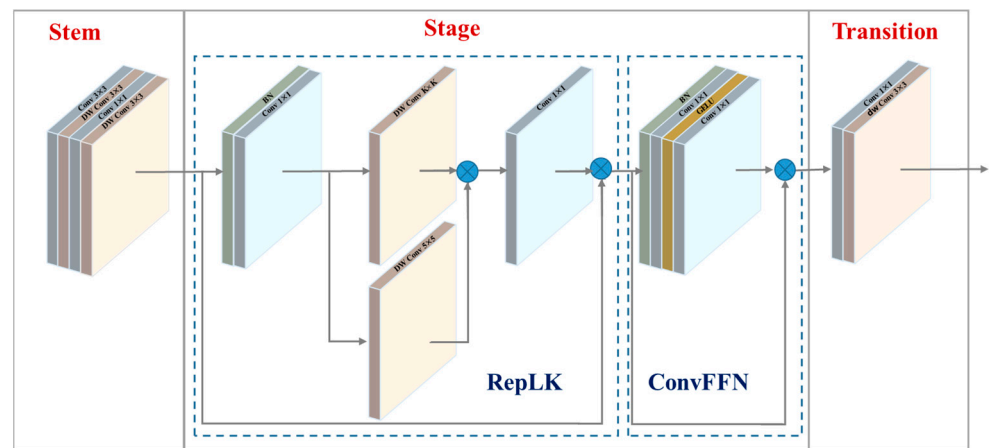
Figure 3. RepLKNet network architecture [30].

2.2.3. CBAM Attention

This study introduces the CBAM to improve the model's feature representation capability. This module integrates both channel and spatial attention mechanisms, providing the model with comprehensive and effective feature extraction capabilities [32]. Figures 4 and 5 depict the core idea of the CBAM, which aims to refine the input feature map in two stages: first, by refining the dependencies among channels using the channel attention module (CAM), and then by enhancing features along the spatial dimension using the spatial attention module (SAM).
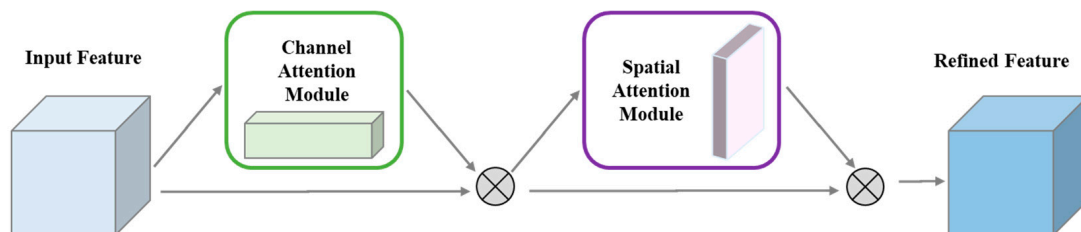

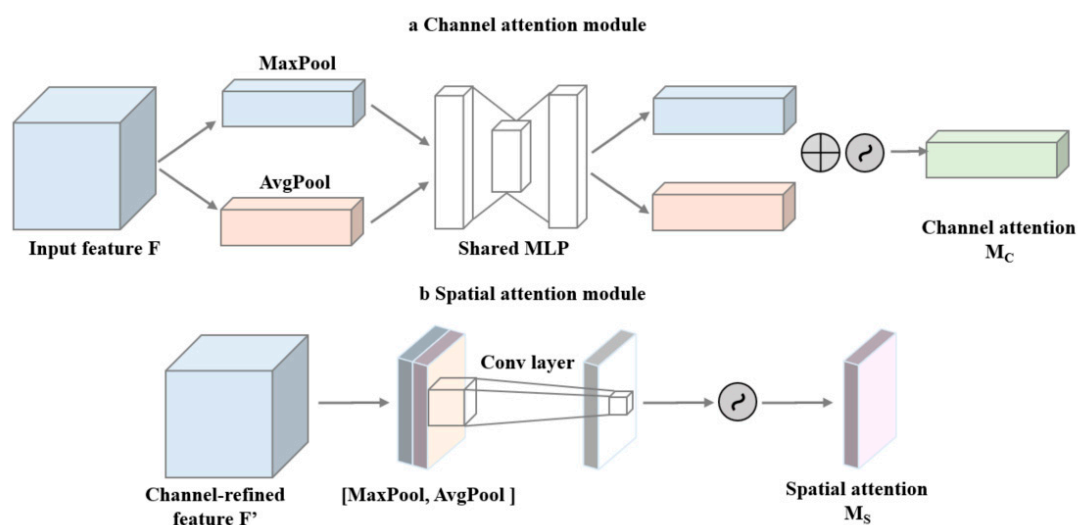**Figure 4.** CBAM architecture [32].

Figure 5. Computational process of the attention sub-module [32].

The CAM assesses the significance of each channel, thereby weighting the channel dimension of the feature map to highlight the contributions of key feature channels. It uses

global max pooling and global average pooling to capture local details and global information, respectively, from the feature map. The output of CAM is articulated as follows:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))), \tag{1}$$

where A and M stand for the global average pooling and max pooling procedures, respectively, and *F* is the input feature map. Additionally, the multi-layer perceptron is represented by M, and the sigmoid activation function is represented by σ.

The SAM further optimizes the spatial dimension of the feature map by generating a spatial attention map along the channel dimension using pooling operations. The significance of various geographical places in the feature map is reflected in this map. The model adaptively improves the feature responses of significant regions while reducing noisy or irrelevant parts, by multiplying the spatial attention map by the feature map. The output of the SAM is articulated as follows:

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right), \tag{2}$$

where [*AvgPool* (*F*); *MaxPool* (*F*)] is the concatenation of the average pooling and the max pooling results along the channel axis, and $f^{7\times7}$ indicates a 7×7 convolution operation. This two-stage attention mechanism allows the CBAM to effectively capture essential information in the feature map's channel and spatial dimensions, thereby enhancing the model's ability to perceive key features.

2.2.4. DynamicHead

This study introduces a dynamic detection head to integrate multi-dimensional attention into the original YOLOv11 detection head mechanism. The DynamicHead aims to unify scale, spatial, and task awareness during object detection by incorporating various attention mechanisms, thereby considerably improving the representational capacity of the detection head while maintaining computational efficiency [33]. Figure 6 illustrates the DynamicHead architecture's three attention mechanisms. A 3D tensor is created by first extracting feature pyramids using different backbone networks, scaling them to the same scale, and then feeding the result into the DynamicHead. Then, many DyHead blocks with task-aware, scale-aware, and spatial-aware attention mechanisms are placed one after the other.
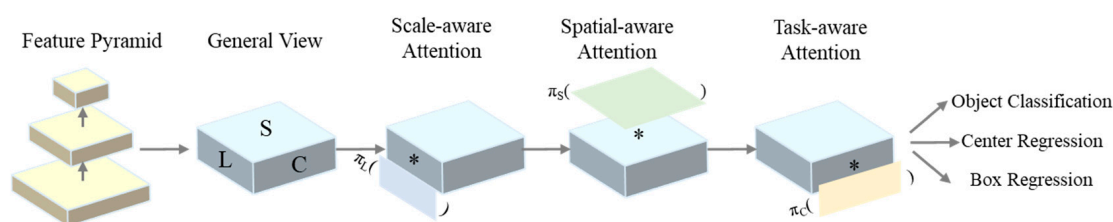

Figure 6. Illustration of the DynamicHead approach [33].

Scale-aware attention, one of these processes, is used at the feature level by dynamically combining information based on the semantic significance of various scales. This improves the detection head's capacity to handle items of different sizes within the picture by increasing the sensitivity of the feature map to scale fluctuations affecting the foreground objects. The spatial dimension (height × breadth) is where spatially aware attention is applied. In order to concentrate on discriminative areas that are consistently present across spatial locations and feature levels, this technique first uses deformable convolution to sparsify the attention learning process before aggregating cross-level data at the same

spatial location. The feature map becomes sparser and more focused on the discriminative spatial locations of foreground objects as a result of this adaptive aggregation of several feature levels.

These attention mechanisms operate in different dimensions, complementing each other and collectively enhancing the representational capability of the object detection head. Integrating them into a unified framework allows the DynamicHead to effectively address challenges related to scale and space in object detection tasks, thereby improving detection performance.

2.2.5. Wise-IoU

This approach may result in a distorted evaluation of the outcomes [34]. However, the Weighted Intersection over Union function (WIoU) mitigates this issue by incorporating the areas between the predicted and ground truth bounding boxes into the IoU computation, thus reducing potential biases present in conventional IoU assessments [35]. To calculate the IoU score between the predicted and ground truth bounding boxes, the spatial interaction between them is examined. More specifically, the distance between the centers of the predicted and ground truth bounding boxes is measured and used as a benchmark for the maximum feasible distance between the two. Based on the regions between the two boxes, a weighting coefficient is calculated, measuring the relationship between the two boxes, and this is used to weight the IoU score. By introducing the regions between the boxes and the weighting coefficient, WIoU can more accurately evaluate object detection results, avoiding biases inherent in traditional IoU.

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2\left(b^A, b^B\right)}{c^2} + \alpha v \tag{3}$$

$$L_{\text{WIoU}} = r R_{\text{WIoU}} L_{\text{IoU}}, r = \frac{\beta}{\delta \alpha^{\beta - \alpha}} \tag{4}$$

$$\beta = \frac{L_{\text{IoU}}^*}{L_{\text{IoU}}} \in [0, +\infty) \tag{5}$$

$$R_{\text{WIoU}} = \exp\left(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{\left(c_w^2 + c_h^2\right)^*}\right) \tag{6}$$

where L_{CIoU} is an enhanced Intersection over Union (IoU) loss function that accounts for the impact of both the distance between bounding box centers and their aspect ratios. L_{WIoU} represents a weighted IoU loss function, which modifies the traditional IoU loss by employing a weighting coefficient r alongside the region weight R_{WIoU}. Additionally, parameters α and v serve to balance considerations related to center distances and aspect ratios.

The R_{WIoU} is the region weight based on the distance between the centers of the predicted and ground truth bounding boxes, and β is a dynamically adjusted weighting coefficient that balances the IoU loss across different samples. By introducing region weights and a dynamic adjustment mechanism, WIoU can precisely evaluate the object detection results, enhancing the model's detection performance.

2.2.6. DynamicATSS

Label assignment is crucial in modern object detection models, as different label assignment strategies affect performance outcomes [36].

This paper employs a straightforward yet highly effective dynamic label assignment strategy called Dynamic ATSS, which incorporates predictions into the label assignment process for anchors (Figure 7) [37]. In the initial stages of training, predictions tend to be inaccurate owing to random initialization. Therefore, as in previous methods, the anchors

are defined by label indications. However, as training progresses and predictions improve, the predictions progressively take precedence over the combined Intersection over Union (IoU), leading to more refined label assignments. This strategy provides dynamic label assignment determined by the training state and the corresponding predictions.
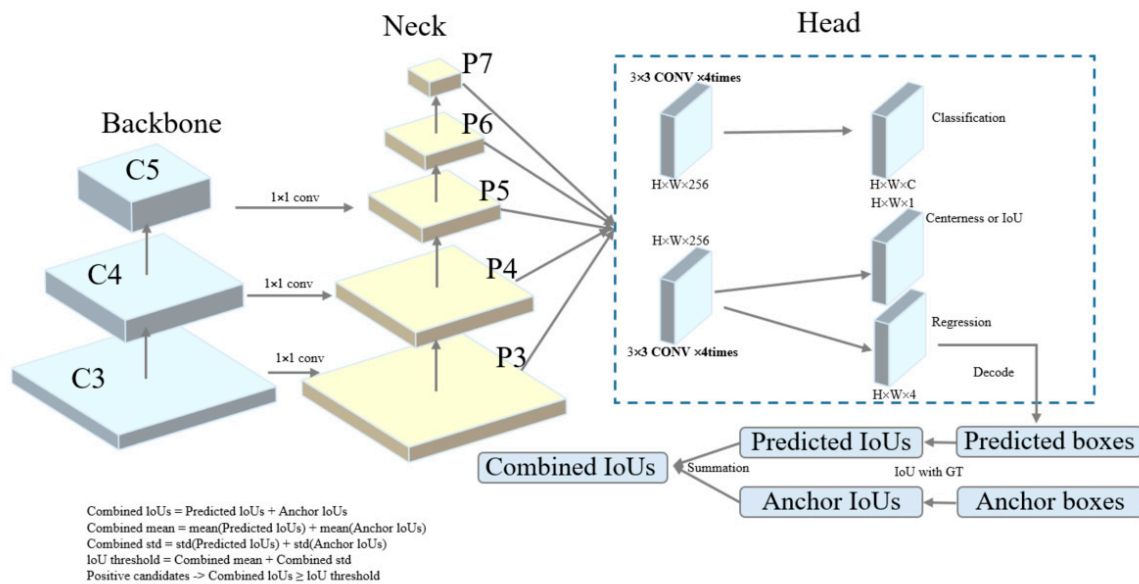

Figure 7. Illustration of the DynamicATSS approach [37].

2.3. Model Training and Evaluation Metrics

2.3.1. Maize Leaf Algorithm Model Training Environment

We utilized YOLOv11 as the base model for our research. YOLOv11 is publicly available through the Ultralytics GitHub repository "(https://github.com/ultralytics/ultralytics (accessed on 30 December 2024)", ensuring reproducibility and accessibility for the research community. The experimental setup in this study comprised a 64-bit Ubuntu Server 22.04 LTS operating system, utilizing an NVIDIA RTX 4090D GPU with 24 GB of video memory and 80 GB of host memory. The programming language was Python 3.10, with GPU acceleration enabled using CUDA v11.8. The training was performed using the deep learning framework PyTorch 2.1.2. Table 1 summarizes the training parameter settings used in our study. These parameters were carefully chosen through preliminary experiments and via the literature review to optimize model performance and training efficiency.

Table 1. Training Parameter Settings.

Parameter	Value	Parameter	Value
Epochs	300	Optimizer	SGD
Patience	50	Weight_decay	0.0005
Batch	8	momentum	0.937
Imgsize	640	Warmup_momentum	0.8
Workers	8	Lrf	0.05

2.3.2. Evaluation Metrics

We divided the maize leaf dataset into a training set, validation set, and test set in a 7:2:1 ratio to facilitate model training, optimization, and evaluation for subsequent research. The evaluation metrics employed in the maize leaf disease detection model in this study included precision, recall, F1 score, mean average precision (mAP), detection speed, model parameters, computational load (GFLOPs), and memory cost. Higher precision indicates

more reliable positive predictions by the model. Recall measures the model's ability to detect all actual positives, with higher values indicating fewer false negatives. The F1 score balances precision and recall, providing a comprehensive measure that is especially useful for imbalanced datasets. Finally, mAP evaluates the model's ranking of predictions across multiple classes, with higher values indicating better overall performance. *Precision*, *recall*, *F*1 score, and mAP are formulated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

$$\text{AP} = \int_0^1 Precision \cdot Recall dr \tag{10}$$

The model's real-time performance was assessed based on detection speed, model parameters, computational load (GFLOPs), and memory cost. Detection speed refers to the time required for the model to process an image, measured in seconds (s) or milliseconds (ms). Faster detection speeds indicate better real-time performance in practical application. Model parameters are the total number of trainable parameters, measured in millions (M); fewer parameters generally indicate a more lightweight model that is easier to deploy on resource-constrained devices. GFLOPs refer to the number of floating-point operations required during the model's forward propagation, measured in billions of floating-point operations (gigaFLOPs). Lower computational load indicates lower hardware resource requirements and faster runtime speeds. Memory cost is the memory required for the model to run, measured in megabytes (MB). Lower memory cost indicates that the model is more suitable for running on memory-constrained devices.

3. Results

3.1. Comparative Experiments of Different Backbone Networks

This study included a series of comparative experiments aimed at comprehensively examining the profound impact of diverse backbone networks on the performance of the YOLOv11 model in the detection tasks related to maize leaf diseases. The experiments evaluated five mainstream lightweight backbone network models, including CFNet, FasterNet, GhostNetV2, MobileViTBv3, and RepLKNet, where each was integrated into the YOLOv11 model architecture. These models were then evaluated against YOLOv11's native backbone network, C2K3. Table 2 reports the corresponding experimental results.

Table 2 highlights significant performance variations in the maize disease detection tasks depending on the backbone network. For instance, the YOLOv11 model with RepLKNet as the backbone (YOLOv11s-RepLKNet) demonstrated notable comprehensive advantages in four key aspects: detection accuracy, detection speed, model complexity, and resource usage, particularly excelling in detection accuracy. RepLKNet achieved the highest values in both precision and F1 score, reaching 89.1% and 84.2%, respectively. This indicates that RepLKNet has a significant advantage in terms of detection accuracy, effectively reducing false positives and missed detections. Additionally, its recall reached 79.9%, demonstrating its considerable detection completeness and robust performance. Although its mAP@0.5~0.95 was slightly lower than some of the models, its average precision at different confidence thresholds (mAP@0.5) remained high (87.1%), proving its detection capability under multi-threshold conditions.

Table 2. Maize Disease Detection Performance of Different Backbone Network Models.

Model	Precision/%	Recall/%	F1 Score/%	mAP@0.5/%	mAP@0.5~0.95/%	Detection Speed/ms	Parameter/M	GFLOPs	MemoryCost/MB
YOLOv11s-C2K3	87.7	78.0	82.6	85.3	63.5	1.3	10.52	21.2	19.2
YOLOv11s-CFNet	87.3	79.8	83.4	85.6	63.9	3.2	8.89	22.7	18.1
YOLOv11s-FasterNet	86.5	80.2	83.2	85.5	63.3	2.4	9.04	23.7	18.5
YOLOv11s-GhostNetV2	87.3	77.1	81.9	84.6	62.6	3.3	8.57	21.2	17.6
YOLOv11s-Mobile-ViTBv3	83.2	80.9	82.0	85.8	63.5	4.4	10.60	34.8	21.6
YOLOv11s-RepLKNet	89.1	79.9	84.2	87.1	66.7	3.2	9.85	25.2	20.2

The detection speed of YOLOv11s-RepLKNet was 3.2 ms per image, comparable to YOLOv11s-CFNet and significantly better than YOLOv11s-MobileViTBv3 (4.4 ms per image). This demonstrates that RepLKNet can achieve fast detection speeds while maintaining high accuracy, rendering it appropriate for real-time detection of maize diseases in practical application scenarios.

From the perspective of model complexity and resource usage, although YOLOv11s-RepLKNet has a slightly higher parameter count and computational load compared with some lightweight models, the significant improvement in its detection accuracy justifies its complexity with reasonable cost-effectiveness. In contrast, other models like YOLOv11s-C2K3, while faster in detection speed, have slightly inferior detection accuracy. Through its unique architectural design, RepLKNet balances lightweight design and high performance, enabling it to excel in complex disease detection tasks. The memory footprint of YOLOv11s-RepLKNet is 20.2 MB; although slightly higher than some lightweight models, this is still within a reasonable range. Considering its advantages in detection accuracy and speed, this level of resource usage is acceptable and can meet the demands of most practical application scenarios.

In summary, the YOLOv11 model with RepLKNet as the backbone performs exceptionally well in maize disease detection tasks, particularly in detection accuracy. Its outstanding performance in key metrics such as precision and F1 score, combined with a reasonable balance between detection speed and resource usage, makes it a highly competitive choice for these tasks. These results fully demonstrate RepLKNet's excellent trade-off between lightweight design and high-performance detection, delivering a reliable and efficient solution for the detection of maize diseases.

3.2. Comparative Experiments of Various Attention Mechanisms

The subsequent trial examined the influence of various attention mechanisms on the performance of the YOLOv11 model for detecting maize diseases. The trials challenged six common attention mechanisms, including CBAM, EC, EMA, GAM, SA, SimAM, and SK; each was integrated into the YOLOv11 model for performance evaluation. The proposed YOLOv11-RCDWD model in this study incorporates the CBAM attention mechanism within the YOLOv11 model. The corresponding experimental results are presented in Table 3. The CBAM attention mechanism demonstrated excellent performance across multiple metrics, particularly excelling in precision (90.5%), F1 score (84.1%), mAP@0.5 (86.2%), and memory consumption (17.3 MB). Although the detection speed slightly decreased from 1.3 ms to 1.6 ms, the model's parameter count (8.49 M) was reduced by approximately 2.03 M compared with the baseline model's 10.52 M. Additionally, its computational load (20.6 GFLOPs) is slightly lower than the baseline model's 21.2 GFLOPs, giving it a clear ad-

vantage in overall performance. Therefore, CBAM is the most suitable attention mechanism for maize disease detection tasks.

Table 3. Results Using Different Attention Mechanisms.

Model	Precision/%	Recall/%	F1 Score/%	mAP@0.5/%	mAP@0.5~0.95/%	Detection Speed/ms	Parameter/M	GFLOPs	MemoryCost/MB
YOLOv11s	87.7	78.0	82.6	85.3	63.5	1.3	10.52	21.2	19.2
YOLOv11s-CBAM	90.5	79.4	84.1	86.2	64.6	1.6	8.49	20.6	17.3
YOLOv11s-EC	87.7	78	82.6	85.3	63.5	4.2	9.41	21.3	19.2
YOLOv11s-EMA	86.4	81.1	83.7	86.9	64.6	3.3	9.08	21.7	18.5
YOLOv11s-GAM	88.7	76.6	82.2	84.1	62.4	4.6	11.91	23.3	24.2
YOLOv11s-SA	84.8	81.4	83.1	86.6	64.5	4.1	8.36	20.2	17.2
YOLOv11s-SimAM	85.5	78.8	82.0	86	63.7	3.7	8.42	20.5	17.5

The EMA attention mechanism performed well in recall (81.1%) and mAP@0.5~0.95 (64.6%) but with disadvantages in detection speed (3.3 ms) and parameter count (9.08 M), making it suitable for tasks with high recall requirements. The GAM attention mechanism performed well in precision (88.7%) and F1 score (82.2%), although its parameter count (11.91 M), computational load (23.3 GFLOPs), and memory consumption (24.2 MB) were found to be relatively high and its detection speed is slower (4.6 ms), making it suitable for tasks less sensitive to model complexity. The SA and SimAM attention mechanisms performed moderately across multiple metrics, particularly in precision (SA: 84.8%, SimAM: 85.5%) and F1 score (SA: 83.1%, SimAM: 82.0%), indicating they may not be suitable for maize disease detection tasks.

In summary, the YOLOv11s-CBAM model performs the best in maize leaf disease detection tasks, maintaining high detection accuracy while effectively reducing model parameter count and memory consumption, rendering it appropriate for implementation in real-world applications.

3.3. Ablation Experiments

We also conducted ablation studies by systematically adding various modules to the baseline model, e.g., the RepLKNet network module, CBAM attention mechanism module, DynamicHead, WIoU loss function, and DynamicATSS sample allocation strategy, to assess their impacts on the overall performance of YOLOv11-RCDWD. To maintain the validity of comparisons, all models were trained on a consistent dataset under standardized parameter settings. The comprehensive results of the ablation studies are detailed in Table 4.

Table 4. Ablation Experiment Results.

RepLKNet	CBAM	DynamicHead	WIoU	DynamicATSS	Precision/%	Recall/%	F1 Score/%	mAP@0.5/%	mAP@0.5~0.95/%	Detection Speed/ms	Parameter/M	GFLOPs	MemoryCost/MB
-	-	-	-	-	87.7	78.0	82.6	85.3	63.5	1.3	10.52	21.2	19.2
√	-	-	-	-	89.1	79.9	84.2	87.1	66.7	3.2	9.85	25.2	20.2
-	√	-	-	-	90.5	79.4	84.1	86.2	64.6	1.6	8.49	20.6	17.3
-	-	√	-	-	87.8	82.1	83.8	85.5	65.0	6.5	9.72	21.2	19.8
-	-	-	√	-	86.6	81.0	83.7	86.7	64.4	1.2	10.2	21.3	19.2
-	-	-	-	√	86.3	80.1	83.1	85.6	63.5	3.8	10.3	21.3	19.2
√	√	√	√	√	92.6	85.4	88.9	90.2	72.5	1.6	9.41	19.3	16.4

Introducing the RepLKNet module significantly improved the model's detection accuracy. When used alone, RepLKNet increased the precision from 87.7% to 89.1%, recall from 78.0% to 79.9%, mAP@0.5 from 85.3% to 87.1%, and mAP@0.5~0.95 from 63.5% to 66.7%. However, the detection speed increased from 1.3 ms to 3.2 ms, and GFLOPs

increased from 21.2 M to 25.2 M, indicating that while RepLKNet enhanced performance, it also significantly increased the computational burden.

In contrast, the CBAM attention mechanism module improved the precision to 90.5%, while significantly reducing the number of parameters (from 10.52 M to 8.49 M) and memory overhead (from 19.2 MB to 17.3 MB). Moreover, the detection speed slightly increased to 1.6 ms, demonstrating a good balance between accuracy and efficiency.

The DynamicHead module remarkably improved the model's recall, which increased from 78.0% to 82.1%. However, the detection speed significantly decreased to 6.5 ms, with slight increases in parameter values and memory overhead. This suggests that DynamicHead is more suitable for scenarios where high recall is required but may not be appropriate for tasks with strict real-time requirements.

The WIoU loss function and DynamicATSS sample allocation strategy had relatively limited impacts on model performance. While WIoU slightly improved mAP@0.5 to 86.7% and mAP@0.5~0.95 to 64.4% while maintaining a high detection speed of 1.2 ms, DynamicATSS significantly increased the detection time to 3.8 ms with no significant performance improvement.

Ultimately, when all these modules were combined, the model achieved optimal comprehensive performance. Indeed, the precision, recall, F1 score, and mAP@0.5 reached 92.6%, 85.4%, 88.9%, and 90.2%, respectively, and mAP@0.5~0.95 increased to 72.5%. Meanwhile, the detection speed remained at 1.6 ms, the number of parameters was reduced to 9.41 M, GFLOPs decreased to 19.3 M, and memory overhead was reduced to 16.4 MB. This indicates that the synergistic action of these modules significantly enhanced detection accuracy and optimized computational efficiency and memory usage, offering higher practicality and deployability for real-world applications.

In summary, the RepLKNet and CBAM modules improved the model's precision and detection accuracy. At the same time, the DynamicHead and DynamicATSS strategies enhanced the model's comprehensive performance by optimizing recall and sample allocation. The WIoU loss function also improved the model's localization accuracy and generalization ability without increasing computational complexity. Integrating these modules enabled the YOLOv11-RCDWD model to perform exceptionally well in maize leaf disease detection while successfully balancing detection speed and model complexity.

3.4. Comparative Experiments of the Performance of Different Network Models

To validate the effectiveness of YOLOv11s-RCDWD, the proposed model was tested against several classic object detection models, including Faster R-CNN [38], SSD [39], YOLOv5s [40], YOLOv6s [41], YOLOv7 [42,43], YOLOv9s [44], YOLOv9c [45], YOLOv10s [46], and YOLOv11s, under the same training environment. To ensure the models were optimized specifically for our target dataset, we trained them from scratch using the dataset specified in this study. The evaluation metrics were precision, recall, F1 score, mAP@0.5, mAP@0.5~0.95, detection speed, parameter count, GFLOPs, and memory cost. The corresponding experimental results are reported in Table 5.

The improved model, YOLOv11s-RCDWD, significantly outperformed the competitor models in all the accuracy metrics. Its precision reached 92.6%, recall was 85.4%, and the F1 score was 88.9%, notably higher than Faster R-CNN or SSD. Although the YOLO series models from YOLOv5s to YOLOv11s showed continuous improvements in accuracy, they all fell short of the improved model YOLOv11s-RCDWD. Particularly in the mAP@0.5 and mAP@0.5~0.95 metrics, YOLOv11s-RCDWD achieved 90.2% and 72.5%, respectively, representing improvements of 4.9% and 9.0% over YOLOv11s, indicating its stronger capability for object detection in complex scenarios.

Table 5. Results of Maize Disease Detection Performance Among Different Network Models.

Model	Precision/%	Recall/%	F1 Score/%	mAP@0.5/%	mAP@0.5 ~0.95/%	Detection Speed/ms	Parameter/M	GFLOPs	MemoryCost/MB
Faster R-CNN	82.6	77.2	80.0	80.2	60.1	23.0	28.86	48.6	42.3
SSD	85.2	78.1	80.9	82.1	61.6	15.2	25.6	36.2	30.6
YOLOv5s	85.9	78.9	82.3	84.3	61.0	3.2	7.81	18.7	16.0
YOLOv6s	82.8	77.9	80.3	81.9	60.5	2.5	15.97	42.8	32.3
YOLOv7	87.7	79.6	83.5	86	63.1	3.9	9.82	23.4	20.0
YOLOv9s	88.6	83.3	85.9	88.7	65.4	2.7	6.31	22.6	13.3
YOLOv9c	87.6	83.3	85.4	88.0	67.0	4.2	21.35	84.0	43.3
YOLOv10s	88.6	78.2	83.1	85.1	63.4	2.6	8.03	24.5	16.6
YOLOv11s	87.7	78.0	82.6	85.3	63.5	1.3	10.52	21.2	19.2
YOLOv11s-RCDWD	92.6	85.4	88.9	90.2	72.5	1.6	9.41	19.3	16.4

Regarding parameter count and GFLOPs, YOLOv11s-RCDWD remained largely consistent with YOLOv11s, at 9.41 M and 10.52 M, respectively. These values are significantly lower than those achieved by more complex models like YOLOv9c (21.35 M parameters and 84.0 GFLOPs), whose memory cost was 16.4 MB, less than YOLOv11s. The improved YOLOv11s-RCDWD demonstrated superior efficiency in resource utilization.

The YOLOv11s-RCDWD model demonstrates improved equilibrium among accuracy, speed, and complexity. Notably, its high accuracy and efficiency make it particularly suitable for practical agricultural scenarios such as maize disease detection tasks, and it offers significant advantages when deployed on resource-constrained mobile devices. The YOLOv11s-RCDWD model excels in maize disease detection tasks, with both accuracy and efficiency surpassing other comparative models. The improved model significantly enhances detection performance by introducing RepLKNet, CBAM, DynamicHead, WIoU loss function, and the DynamicATSS sample allocation strategy, providing an effective solution for intelligent detection of maize leaf diseases in complex agricultural environments.

3.5. Visualization of Analysis Results

This study comprehensively evaluated and analyzed the performance of the YOLOv11s-RCDWD model in corn leaf disease identification, using a confusion matrix. The confusion matrix visually presents the model's classification performance across various categories, including correct and incorrect classifications. In Figure 8, the diagonal elements of the confusion matrix indicate the number of samples accurately classified by the model. In contrast, the off-diagonal elements represent instances of misclassification among different categories. Detailed analysis of the confusion matrix analysis enables a better understanding of the model's recognition capabilities and shortcomings across different disease categories.

The experimental results indicate that the improved YOLOv11s-RCDWD model achieved generally high accuracy in identifying different disease categories, particularly in recognizing healthy leaves, for which its accuracy rate was 96.82%. Its recognition accuracy for northern leaf blight and gray leaf spot is also relatively high, reaching 93.49% and 92.55%, respectively. However, the model exhibited some confusion in distinguishing between round spot disease and southern leaf blight, with accuracy rates of 89.68% and 90.99%, respectively. This may be attributed to the potential visual similarity between these two diseases, leading to misclassification by the model when differentiating between them.

Figure 9 compares the detection results of the improved YOLOv11s-RCDWD model against the original model. The mutual misclassifications observed among gray leaf spot, round spot disease, and southern leaf blight may stem from the subtle differences present in the early stages of these diseases. Consequently, the model may not have developed sufficiently robust features to effectively distinguish between these conditions, resulting in erroneous classifications.
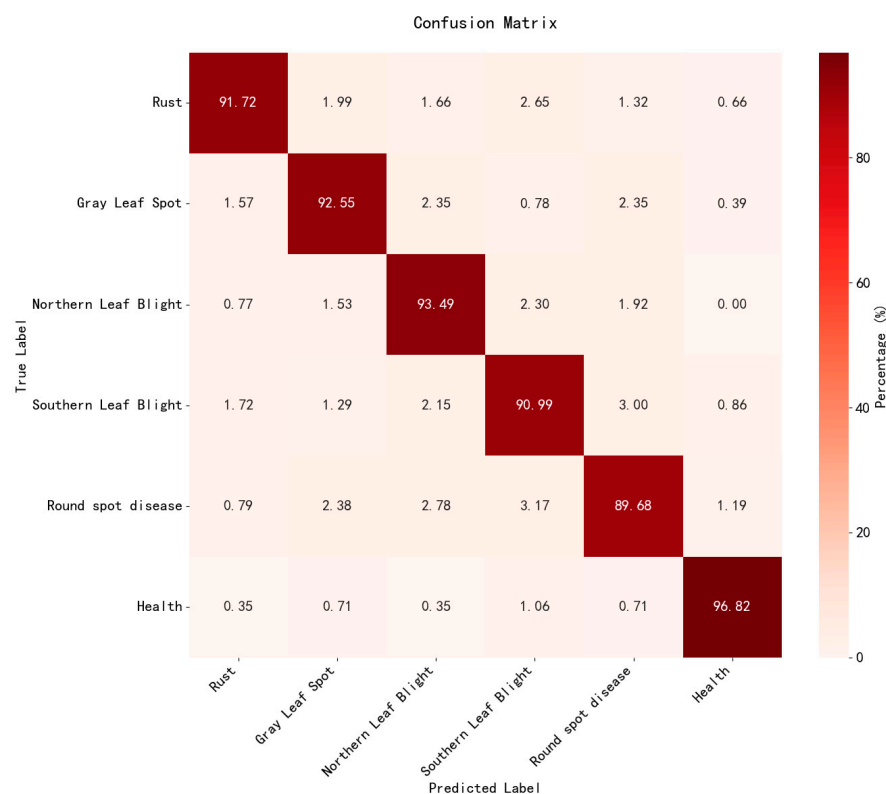
Figure 8. Confusion Matrix: YOLOv11s-RCDWD.

Figure 9 compares the detection results of the improved YOLOv11s-RCDWD model against the original model. The mutual misclassifications observed among gray leaf spot, round spot disease, and southern leaf blight may stem from the subtle differences present in the early stages of these diseases. Consequently, the model may not have developed sufficiently robust features to effectively distinguish between these conditions, resulting in erroneous classifications.
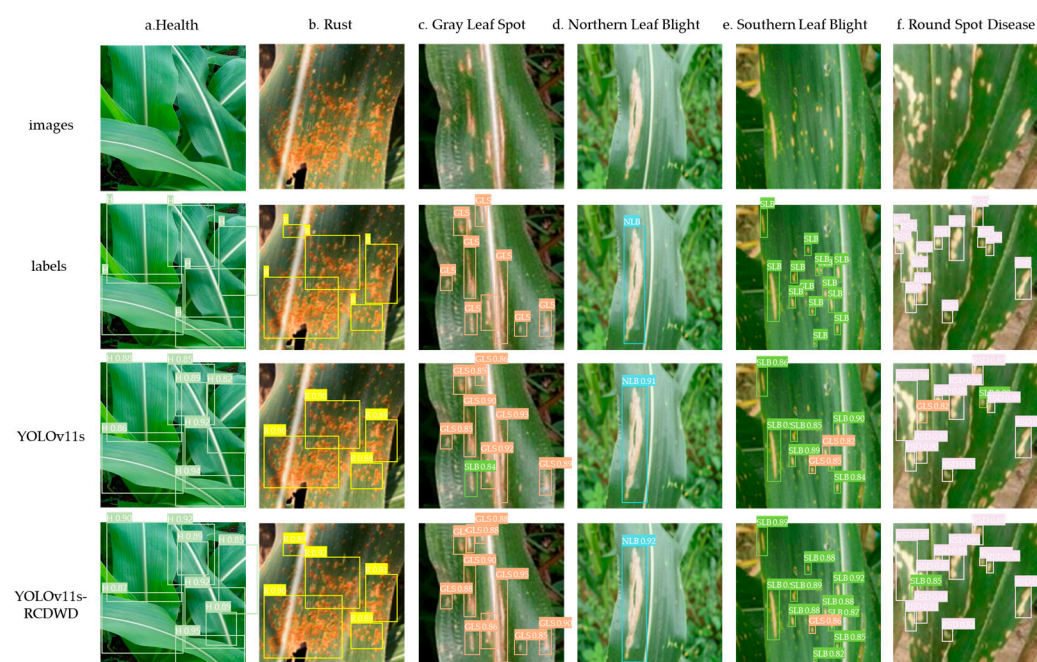
Figure 9. The detection results of the improved YOLOv11s-RCDWD model compared with the original model.

In summary, the YOLOv11s-RCDWD model demonstrated high accuracy in the maize leaf disease identification task, particularly excelling in recognizing healthy leaves and northern leaf blight. However, the model still has certain limitations in distinguishing disease categories with similar visual features, such as rust, gray leaf spot, southern leaf blight, and round spot disease. These limitations may stem from the similar morphological characteristics of the diseases on the leaves or insufficient sample sizes for certain categories in the training data. Thus, there is still room for improvement in the model's performance through optimization and training.

4. Discussion

Existing YOLO-based models have demonstrated notable achievements in various detection tasks, but they still face limitations in complex agricultural scenarios. For example, Li et al. [20] proposed YOLO-Leaf for apple leaf disease detection, which utilized DSConv, BiFormer, and IF-CIoU to enhance feature extraction and bounding box regression. Despite its improved accuracy, YOLO-Leaf still struggles with generalization in diverse environmental conditions. Similarly, Lu et al. [21] developed COTTON-YOLO by introducing the C2F-CBAM module and Gold-YOLO neck architecture, yet this model remains limited by rigid sample allocation strategies. Yang et al. [22] enhanced the detection of maize leaf spot by integrating Slim-neck and GAM attention into YOLOv8, but challenges persist in efficiently handling small and densely packed targets.

To address these limitations, our proposed model introduces novel techniques and optimizations tailored for agricultural detection tasks. The proposed YOLOv11s-RCDWD model demonstrates outstanding performance in detecting maize leaf diseases and pests by incorporating the RepLKNet backbone network, CBAM attention mechanism, Dynamic-Head detection head, WIoU loss function, and DynamicATSS label assignment strategy. Its detection accuracy, speed, and resource efficiency significantly surpass existing models.

The model demonstrated an F1 score of 88.9%, precision of 92.6%, and a recall value of 85.4%. These findings represent improvements of 4.9% and 9.0% over the YOLOv11 baseline model. This suggests that the enhanced model is better able to locate and detect pests and diseases of maize leaves in complicated environments. Notably, the detection time for a single image is only 1.6 ms, significantly faster than traditional models (e.g., Faster R-CNN and YOLOv8), meeting the requirements for real-time detection. Furthermore, the model's parameter count and computational load are significantly lower than those of comparative models, with a memory footprint of only 16.4 MB, making it suitable for deployment on resource-constrained agricultural mobile devices.

Compared with classic models such as SSD, Faster R-CNN, and YOLOv8, YOLOv11s-RCDWD exhibits clear advantages in terms of detection accuracy and speed. For instance, while Faster R-CNN performs well in detecting small targets, its detection speed is slower, making it challenging to meet real-time detection needs. On the other hand, YOLOv8 improves the speed but suffers from lower detection accuracy against complex backgrounds. Compared with recently proposed improved models, YOLOv11s-RCDWD outperforms other approaches in terms of detection accuracy and resource efficiency. For example, although YOLOv9c provides high detection accuracy, its large parameter count and computational load make deployment on devices with limited resources difficult. In contrast, the lightweight design of YOLOv11s-RCDWD significantly reduces the model's complexity. Compared with common attention mechanisms (e.g., ECA, GAM, and SimAM) and loss functions (e.g., CIoU and DIoU), CBAM, and WIoU enhance detection accuracy and precise localization. For instance, the CBAM's spatial attention and dual-channel processes greatly enhance the model's capacity to extract important characteristics. By dynamically

modifying bounding box regression weights, WIoU simultaneously improves the model's learning capacity for difficult data.

There are a number of important reasons why the YOLOv11s-RCDWD model performs better than other variants. By adding large-kernel convolutions, the RepLKNet backbone network improves the model's capacity to collect global characteristics. This architecture is particularly effective for detecting maize leaf diseases and pests in natural scenes characterized by complex backgrounds. Furthermore, the dual-channel and spatial attention processes of the CBAM attention mechanism improve the model's ability to extract features, greatly reducing false positives and missed detections. Additionally, the DynamicHead detection head adapts to detection tasks of varied sizes and complexities by dynamically adjusting its settings, therefore greatly enhancing the model's capacity to recognize multiscale objects. By dynamically modifying regression weights, the WIoU loss function increases bounding box localization accuracy and fortifies the model's capacity to learn from difficult samples. Finally, the DynamicATSS label assignment approach optimizes sample distribution during training by dynamically altering the selection mechanism for positive and negative samples, considerably increasing the model's generalization capabilities.

Despite its excellent performance in detection of maize leaf diseases and pests, the YOLOv11s-RCDWD model has some limitations. The dataset used is limited in size and diversity, affecting the model's generalization ability. Future work should focus on expanding the dataset's scale and diversity to improve the model's robustness in practical applications. Although YOLOv11s-RCDWD demonstrates excellent resource efficiency, there is still room to further reduce computational load and the parameter count. Future research could explore more efficient model compression methods (e.g., knowledge distillation and quantization) and acceleration techniques (e.g., hardware acceleration) to reduce computational costs. Additionally, the current model is primarily designed for detection of maize leaf diseases and pest, and its generalization ability to other crops and scenarios has not been fully validated. Future efforts should extend this method to detection of diseases and pests in different crops, providing more comprehensive technical support for smart agricultural development.

5. Conclusions

This study proposes an improved YOLOv11-based model, YOLOv11-RCDWD, for detecting maize leaf diseases and pests. Incorporating the RepLKNet module, CBAM module, DynamicHead detection head, WIoU loss function, and DynamicATSS label assignment strategy significantly enhances the model's performance in maize leaf disease and pest detection tasks. The experimental results demonstrated the following:

(1) The improved model, YOLOv11s-RCDWD, exhibited the best detection performance, with all accuracy metrics surpassing those of the comparative models. Its precision reached 92.6%, recall was 85.4%, and F1 score was 88.9%, all higher than the other models;

(2) In terms of mAP@0.5 and mAP@0.5~0.95, YOLOv11s-RCDWD achieved 90.2% and 72.5%, respectively, improvements of 4.9% and 9.0% over YOLOv11s, indicating stronger detection capabilities in complex scenarios;

(3) The parameter count and computational load (GFLOPs) of YOLOv11s-RCDWD remained largely consistent with YOLOv11s, at 9.41 M and 10.52 M, respectively, which are significantly lower than those of more complex models like YOLOv9c. Its memory footprint (memory cost) was found to be 16.4 MB, a reduction from YOLOv11s's 19.2 MB, demonstrating the improved model' more efficient utilization of resources;

(4) Different backbone networks exhibited significant performance variations in maize disease detection tasks. The YOLOv11 model with RepLKNet as the backbone

(YOLOv11s-RepLKNet) demonstrated notable comprehensive advantages in four key aspects: detection accuracy, detection speed, model complexity, and resource usage, particularly excelling in detection accuracy. RepLKNet achieved the highest values in both precision and F1 score, reaching 89.1% and 84.2%, respectively;

(5) Comparing seven common attention mechanisms—CBAM, EC, EMA, GAM, SA, SimAM, and SK—demonstrated that the CBAM attention module significantly outperformed other attention mechanisms in terms of detection accuracy. Its precision reached 90.5%, and its F1 score was 84.1%.

Furthermore, YOLOv11s-RCDWD efficiently utilizes computational resources, attaining an ideal equilibrium between detection accuracy and speed. Its streamlined architecture with the fewest parameters and FLOPs ensures minimal usage of computational resources, facilitating deployment on agricultural machinery equipped with embedded systems. The findings indicate that the YOLOv11s-RCDWD model offers substantial advantages in the detection of maize leaf diseases. offering an efficient and practical solution for intelligent detection of agricultural diseases. Overall, YOLOv11-RCDWD performed well in the maize leaf disease and pest detection tasks, offering a robust technical method for the intelligent detection of agricultural diseases. Future research could further optimize the model's computational efficiency and extend its application to broader agricultural scenarios. Additionally, we will further enhance the scale and diversity of the dataset to improve the model's generalization capabilities. We will also investigate more efficient techniques for model compression to satisfy the deployment requirements of mobile devices. Furthermore, we aim to extend this approach to disease and pest detection in various crops, thereby providing more comprehensive technical support for the advancement of smart agriculture.

References

1. Ranum, P.; Peña Rosas, J.P.; Garcia Casal, M.N. Global maize production, utilization, and consumption. *Ann. N. Y. Acad. Sci.* **2014**, *1312*, 105–112. [CrossRef] [PubMed]
2. Ren, L.; Li, C.; Yang, G.; Zhao, D.; Zhang, C.; Xu, B.; Feng, H.; Chen, Z.; Lin, Z.; Yang, H. The Detection of Maize Seedling Quality from UAV Images Based on Deep Learning and Voronoi Diagram Algorithms. *Remote Sens.* **2024**, *16*, 3548. [CrossRef]
3. Savary, S.; Ficke, A.; Aubertot, J.; Hollier, C. Crop losses due to diseases and their implications for global food production losses and food security. *Food Secur.* **2012**, *4*, 519–537. [CrossRef]
4. John, M.A.; Bankole, I.; Ajayi-Moses, O.; Ijila, T.; Jeje, T.; Lalit, P. Relevance of advanced plant disease detection techniques in disease and Pest Management for Ensuring Food Security and Their Implication: A review. *Am. J. Plant Sci.* **2023**, *14*, 1260–1295. [CrossRef]

5. DeChant, C.; Wiesner-Hanks, T.; Chen, S.; Stewart, E.L.; Yosinski, J.; Gore, M.A.; Nelson, R.J.; Lipson, H. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* **2017**, *107*, 1426–1432. [CrossRef] [PubMed]

6. Setiawan, W.; Rochman, E.; Satoto, B.D.; Rachmad, A. Machine learning and deep learning for maize leaf disease classification: A review. *J. Phys. Conf. Ser.* **2022**, *2406*, 12019. [CrossRef]

7. Jafar, A.; Bibi, N.; Naqvi, R.A.; Sadeghi-Niaraki, A.; Jeong, D. Revolutionizing agriculture with artificial intelligence: Plant disease detection methods, applications, and their limitations. *Front. Plant Sci.* **2024**, *15*, 1356260. [CrossRef]

8. Abdullah, H.M.; Mohana, N.T.; Khan, B.M.; Ahmed, S.M.; Hossain, M.; Islam, K.S.; Redoy, M.H.; Ferdush, J.; Bhuiyan, M.; Hossain, M.M. Present and future scopes and challenges of plant pest and disease (P&D) monitoring: Remote sensing, image processing, and artificial intelligence perspectives. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 100996.

9. Shi, Y.; Duan, Z.; Qing, S.; Zhao, L.; Wang, F.; Yuwen, X. YOLOV9S-Pear: A lightweight YOLOV9S-Based improved model for young Red Pear Small-Target recognition. *Agronomy* **2024**, *14*, 2086. [CrossRef]

10. Panigrahi, K.P.; Das, H.; Sahoo, A.K.; Moharana, S.C. Maize leaf disease detection and classification using machine learning algorithms. In *Progress in Computing, Analytics and Networking—Proceedings of the ICCAN 2019, Bhubaneswar, India, 14–15 December 2019*; Springer: Singapore, 2020; pp. 659–669.

11. Paul, H.; Udayangani, H.; Umesha, K.; Lankasena, N.; Liyanage, C.; Thambugala, K. Maize leaf disease detection using convolutional neural network: A mobile application based on pre-trained VGG16 architecture. *N. Z. J. Crop Hortic. Sci.* **2024**, *53*, 367–383. [CrossRef]

12. Reddy, J.; Niu, H.; Scott, J.L.L.; Bhandari, M.; Landivar, J.A.; Bednarz, C.W.; Duffield, N. Cotton Yield Prediction via UAV-Based Cotton Boll Image Segmentation Using YOLO Model and Segment Anything Model (SAM). *Remote Sens.* **2024**, *16*, 4346. [CrossRef]

13. Song, Y.; Yang, L.; Li, S.; Yang, X.; Ma, C.; Huang, Y.; Hussain, A. Improved YOLOv8 Model for Phenotype Detection of Horticultural Seedling Growth Based on Digital Cousin. *Agriculture* **2024**, *15*, 28. [CrossRef]

14. 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.* **2021**, *8*, 27–51. [CrossRef]

15. Sharma, A.; Kumar, V.; Longchamps, L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. *Smart Agric. Technol.* **2024**, *9*, 100648. [CrossRef]

16. Zhang, Z.; Yang, Y.; Xu, X.; Liu, L.; Yue, J.; Ding, R.; Lu, Y.; Liu, J.; Qiao, H. GVC-YOLO: A Lightweight Real-Time Detection Method for Cotton Aphid-Damaged Leaves Based on Edge Computing. *Remote Sens.* **2024**, *16*, 3046. [CrossRef]

17. Meng, Z.; Du, X.; Sapkota, R.; Ma, Z.; Cheng, H. YOLOv10-pose and YOLOv9-pose: Real-time strawberry stalk pose detection models. *Comput. Ind.* **2025**, *165*, 104231. [CrossRef]

18. Fan, Y.; Zhang, S.; Feng, K.; Qian, K.; Wang, Y.; Qin, S. Strawberry maturity recognition algorithm combining dark channel enhancement and YOLOv5. *Sensors* **2022**, *22*, 419. [CrossRef]

19. Wang, S.; Zhao, J.; Cai, Y.; Li, Y.; Qi, X.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. A method for small-sized wheat seedlings detection: From annotation mode to model construction. *Plant Methods* **2024**, *20*, 15. [CrossRef]

20. Li, T.; Zhang, L.; Lin, J. Precision agriculture with YOLO-Leaf: Advanced methods for detecting apple leaf diseases. *Front. Plant Sci.* **2024**, *15*, 1452502. [CrossRef]

21. Lu, Z.; Han, B.; Dong, L.; Zhang, J. COTTON-YOLO: Enhancing Cotton Boll Detection and Counting in Complex Environmental Conditions Using an Advanced YOLO Model. *Appl. Sci.* **2024**, *14*, 6650. [CrossRef]

22. Yang, S.; Yao, J.; Teng, G. Corn leaf spot disease recognition based on improved YOLOv8. *Agriculture* **2024**, *14*, 666. [CrossRef]

23. Zhang, X.; Qiao, Y.; Meng, F.; Fan, C.; Zhang, M. Identification of maize leaf diseases using improved deep convolutional neural networks. *IEEE Access* **2018**, *6*, 30370–30377. [CrossRef]

24. Liu, J.; He, C.; Jiang, Y.; Wang, M.; Ye, Z.; He, M. A High-Precision Identification Method for Maize Leaf Diseases and Pests Based on LFMNet under Complex Backgrounds. *Plants* **2024**, *13*, 1827. [CrossRef]

25. Sun, J.; Yang, Y.; He, X.; Wu, X. Northern maize leaf blight detection under complex field environment based on deep learning. *IEEE Access* **2020**, *8*, 33679–33688. [CrossRef]

26. Zhang, J.; Meng, Y.; Yu, X.; Bi, H.; Chen, Z.; Li, H.; Yang, R.; Tian, J. Mbab-yolo: A modified lightweight architecture for real-time small target detection. *IEEE Access* **2023**, *11*, 78384–78401. [CrossRef]

27. Hughes, D.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.

28. Moschidis, C.; Vrochidou, E.; Papakostas, G.A. Annotation tools for computer vision tasks. In Proceedings of the Seventeenth International Conference on Machine Vision (ICMV 2024), Edinburgh, UK, 10–13 October 2024; Volume 13517, pp. 372–379.

29. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725.

30. Ding, X.; Zhang, X.; Han, J.; Ding, G. Scaling up your kernels to 31 × 31: Revisiting large kernel design in cnns. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11963–11975.

31. Luo, W.; Li, Y.; Urtasun, R.; Zemel, R. Understanding the effective receptive field in deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 4905–4913.

32. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

33. Dai, X.; Chen, Y.; Xiao, B.; Chen, D.; Liu, M.; Yuan, L.; Zhang, L. Dynamic head: Unifying object detection heads with attentions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7373–7382.

34. Nowozin, S. Optimal decisions from probabilistic models: The intersection-over-union case. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 548–555.

35. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. *arXiv* **2023**, arXiv:2301.10051.

36. Zhang, F.; Zhou, S.; Wang, Y.; Wang, X.; Hou, Y. Label assignment matters: A gaussian assignment strategy for tiny object detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–12. [CrossRef]

37. Zhang, T.; Luo, B.; Sharda, A.; Wang, G. Dynamic label assignment for object detection by combining predicted ious and anchor ious. *J. Imaging* **2022**, *8*, 193. [CrossRef] [PubMed]

38. 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.* **2016**, *39*, 1137–1149. [CrossRef]

39. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016—Proceedings of the 14th European Conference, Proceedings, Part I 14, Amsterdam, The Netherlands, 11–14 October 2016*; Springer: Cham, Switzerland, 2016; pp. 21–37.

40. Ma, L.; Yu, Q.; Yu, H.; Zhang, J. Maize leaf disease identification based on yolov5n algorithm incorporating attention mechanism. *Agronomy* **2023**, *13*, 521. [CrossRef]

41. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.

42. Wang, C.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.

43. Zhang, C.; Hu, Z.; Xu, L.; Zhao, Y. A YOLOv7 incorporating the Adan optimizer based corn pests identification method. *Front. Plant Sci.* **2023**, *14*, 1174556. [CrossRef]

44. Wang, C.; Yeh, I.; Mark Liao, H. Yolov9: Learning what you want to learn using programmable gradient information. In *Computer Vision—ECCV 2024—Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024*; Springer: Cham, Switzerland, 2024; pp. 1–21.

45. Gharat, K.; Jogi, H.; Gode, K.; Talele, K.; Kulkarni, S.; Kolekar, M.H. Enhanced Detection of Maize Leaf Blight in Dynamic Field Conditions Using Modified YOLOv9. In Proceedings of the 2024 IEEE Space, Aerospace and Defence Conference (SPACE), Bangalore, India, 22–23 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 140–143.

46. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2025**, *37*, 107984–108011.