agronomy

MDPI

Article

DCF-Yolov8: An Improved Algorithm for Aggregating Low-Level Features to Detect Agricultural Pests and Diseases

Lijuan Zhang [1,2], Gongcheng Ding [1,2], Chaoran Li [3] and Dongming Li [1,*]

[1] College of Internet of Things Engineering, Wuxi University, Wuxi 214105, China; zhanglijuan@ccut.edu.cn (L.Z.); 2202103020@stu.ccut.edu.cn (G.D.)
[2] School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China
[3] School of Information Engineering, Changchun College of Electronic Technology, Changchun 130000, China; lichaoran@cust.edu.cn
* Correspondence: ldm0214@163.com

Abstract: The invasion of agricultural diseases and insect pests is a huge difficulty for the growth of crops. The detection of diseases and pests is a very challenging task. The diversity of diseases and pests in terms of shapes, colors, and sizes, as well as changes in the lighting environment, have a massive impact on the accuracy of the detection results. We improved the C2F module based on DenseBlock and proposed DCF to extract low-level features such as the edge texture of pests and diseases. Through the sensitivity of low-level features to the diversity of pests and diseases, the DCF module can better cope with complex detection tasks and improve the accuracy and robustness of the detection. The complex background environment of pests and diseases and different lighting conditions make the IP102 data set have strong nonlinear characteristics. The Mish activation function is selected to replace the CBS module with the CBM, which can better learn the nonlinear characteristics of the data and effectively solve the problems of gradient disappearance in the algorithm training process. Experiments show that the advanced Yolov8 algorithm has improved. Comparing with Yolov8, our algorithm improves the MAP50 index, Precision index, and Recall index by 2%, 1.3%, and 3.7%. The model in this paper has higher accuracy and versatility.

Keywords: nonlinear characteristics; vanishing gradient; Mish; IP102; edge texture; robustness

Citation: Zhang, L.; Ding, G.; Li, C.; Li, D. DCF-Yolov8: An Improved Algorithm for Aggregating Low-Level Features to Detect Agricultural Pests and Diseases. *Agronomy* **2023**, *13*, 2012. https://doi.org/10.3390/agronomy13082012

Academic Editor: Paul Kwan

Received: 14 July 2023
Revised: 24 July 2023
Accepted: 28 July 2023
Published: 29 July 2023

1. Introduction

Agricultural pest and disease infestation is one of the major causes of decreasing crop yield. The severity and impact of pests and diseases depend on the specific types of pests/diseases and crop species. The outbreak of agricultural pests and diseases not only affects crop productivity [1,2] but also leads to ecological damage due to the use of pesticides.

Research on pest and disease datasets is mainly divided into two categories: classification and detection. Traditional machine learning algorithms such as Support Vector Machines, Random Forests, and k-Nearest Neighbors are widely used for crop image classification, assisting farmers in distinguishing different crop varieties and pest and disease types. Chaudhary et al. [3] proposed an improved Random Forest classifier to address the multi-disease classification problem. The classifier consists of the Random Forest machine learning algorithm, attribute evaluators, and instance filtering methods. It achieved promising classification results on the peanut disease dataset. Singh et al. [4] extracted texture and color features using gradient histograms and other methods during the preprocessing stage. They utilized binary particle swarm optimization to select mixed features and achieved promising classification results using a Random Forest classifier. Panchal et al. [5] employed K-means and HSV (Hue, Saturation, Value) to identify infected parts of leaves and utilized GLCM (Gray-Level Co-occurrence Matrix) for feature extraction. This approach effectively

classifies plant diseases. Meenakshi et al. [6] combined the Random Forest algorithm with the InceptionV3 algorithm. They utilized the Random Forest algorithm to extract regions of rice leaf diseases for classification, while the InceptionV3 algorithm was employed for segmentation and achieved the most accurate feature representation. With the advancement of deep learning technology, it has become increasingly feasible to learn complex feature representations and leverage the strengths of deep learning in various domains such as image, video, and sensor data analysis. Suitable deep learning algorithms can enable precise identification and classification of crop pests and diseases, assisting farmers in determining the type of pests/diseases and implementing appropriate prevention and control measures. Ren et al. [7] proposed a feature reuse residual block for target classification. They enhanced the representation capability by using half of the learning and half of the reused target features. The model achieved excellent performance on the IP102 dataset [8] and demonstrated strong generalization capability. Nanni et al. [9] proposed an automatic classifier based on the fusion of saliency methods and convolutional neural networks. They used three different saliency methods for image preprocessing and created three separate images for each saliency method to train different neural networks. The model achieved the best classification performance on the IP102 dataset. Kasinathan et al. [10] proposed a pest detection algorithm consisting of foreground extraction and contour recognition. This algorithm is used for classifying agricultural pests and diseases in highly complex backgrounds, significantly improving classification accuracy and computational performance. Feng et al. [11] optimized the IP102 dataset and proposed a two-stage insect recognition algorithm, TIR, based on convolutional neural networks. By grouping insects according to their similarity in appearance, the algorithm can better extract deep features, achieving excellent detection performance on the IP102 dataset. Zhang et al. [12] proposed a hybrid ResNet model that utilizes additive and multiplicative combinations of convolutional layers to reduce computational performance. On the IP102 dataset, this model achieved a 2% decrease in accuracy but reduced computational performance consumption by 40% in detection. Zhou et al. [13] achieved promising detection performance on the IP102 dataset with a relatively low parameter count by fusing the squeeze-and-excitation-bottleneck block and the maximum feature expansion block.

Research on the IP102 dataset primarily focuses on object classification, with limited studies on object detection. Traditional algorithms exhibit poor performance in feature extraction, while deep learning-based algorithms often emphasize deep features and parameter quantity. This paper concentrates on low-level features and dataset environmental conditions. The improved model effectively handles the dataset's non-linear characteristics and extracts features such as texture edges from pests and diseases. As a result, it achieves high detection accuracy on the IP102 dataset.

Our contributions are summarized as follows:

- DCF (Low-level Feature Aggregation) module: The DCF (Low-level Feature Aggregation) module proposed in this paper aggregates low-level features from the dataset. In the IP102 dataset, pests and diseases occupy a significant proportion of the images, and each image contains a single category of pests and diseases. The model can better learn the textures, edges, and other features of pests and diseases. Compared with the Yolov8 algorithm, our proposed algorithm achieves higher detection accuracy.
- CBM (Channel-wise Mish) module: By analyzing the pixel distribution histograms of images in the IP102 dataset, we observed strong non-linear characteristics in the dataset. Additionally, the effects of lighting and environmental factors also exhibit non-linear features. To better handle these non-linear characteristics in the dataset, we replaced the CBS (Channel-wise Batch Normalization with Sigmoid) module with the Mish activation function. This replacement effectively resolved the issue of gradient vanishing during training, especially when the model depth is deep.
- Performance improvements: The improved Yolov8 algorithm effectively avoids the problem of gradient vanishing during model training, compared with the original

model, within the same number of training epochs. Our proposed algorithm achieves a significant accuracy improvement of 2%, reaching an accuracy rate of 60.8%.

2. Related Work

2.1. Mosaic Data Augmentation

Mosaic data augmentation randomly selects four training images as input and randomly selects a position as the center point of the mosaic image. The four images are adjusted to the same size using operations such as scaling and cropping, and then concatenated into a mosaic object. The coordinates and sizes of the target bounding boxes are updated based on scaling, cropping, and other operations to ensure their correspondence with the mosaic image. The mosaic image generated by the Mosaic data augmentation is shown in Figure 1.

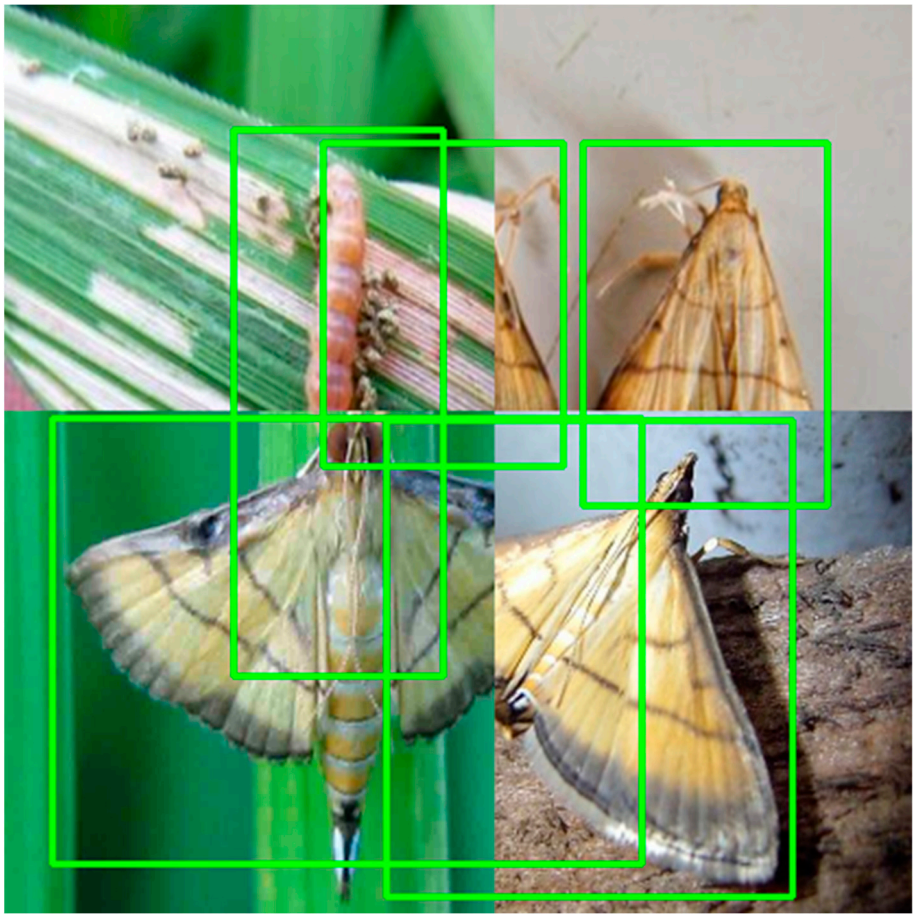


Figure 1. Mosaic data augmentation. The green box is the target box for random cropping and zooming updates.

Mosaic data augmentation enables the generation of a larger variety of samples with different scenes and backgrounds, thereby increasing the diversity of the training data. The mosaic image provides more comprehensive contextual information, allowing the model to extract more accurate contextual cues and improve its generalization capability. The correspondence between the target bounding box coordinates and sizes and the mosaic image enhances the prediction accuracy of the target bounding boxes during the model training process.

2.2. Yolov8

YOLOv8 is a leading-edge and state-of-the-art model that builds upon the improvements and innovations of the previous YOLOv5 [14], which has been widely applied in the agricultural domain [15–18]. YOLOv8 is designed to be fast, accurate, and easy to use, making it suitable for a wide range of tasks including object detection and tracking, instance segmentation, image classification, and pose estimation. Figure 2 illustrates the architecture of the YOLOv8 model.

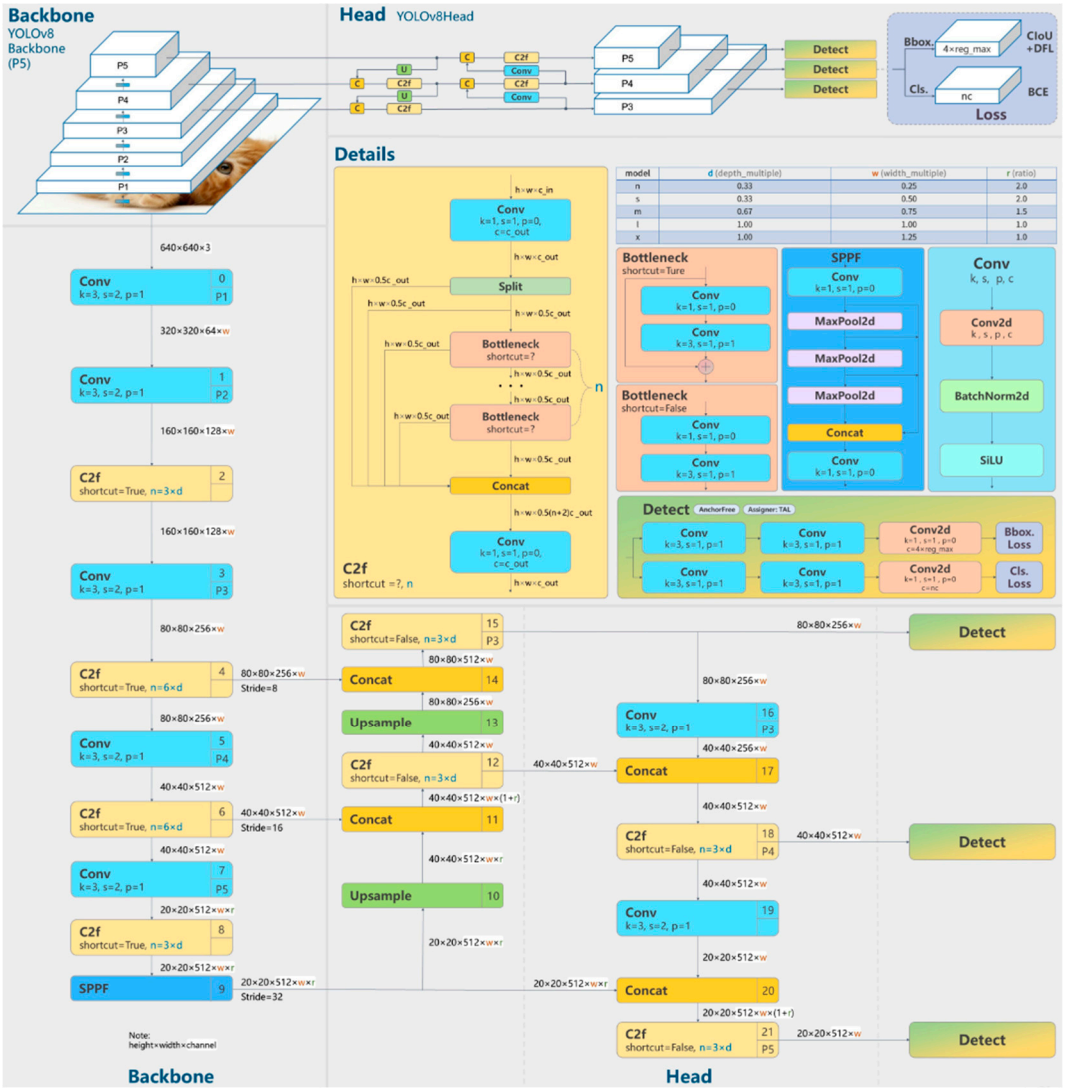


Figure 2. Yolov8 model diagram.

The Darknet-53 [19] backbone network in YOLOv8, inspired by VGG, doubles the number of channels after pooling operations. Additionally, it places 1×1 convolutional kernels between 3×3 convolutional kernels to compress features and employs global average pooling. Batch normalization layers are used to stabilize model training, accelerate convergence, and provide regularization.

The CSP module is a feature extraction module that aims to enhance feature extraction effectiveness by utilizing cross-stage connections. It enables the sharing of features across multiple stages and improves parameter and computational efficiency by partially connecting features from different stages. The Neck part of the YOLOv8 model applies the CSP idea to fuse the original features with features processed through multiple convolution operations, thereby enhancing the feature extraction capability.

The backbone network and Neck part of YOLOv8 draw inspiration from the design principles of YOLOv7 ELAN [20] and introduce the C2F module. The C2F module aims to reduce the parameter count while retaining rich gradient flow information. During the last 10 epochs of training, the Mosaic [21] data augmentation was disabled to improve accuracy.

The Head section of YOLOv8 separates the classification and detection heads. The classification branch continues to use Binary Cross-Entropy (BCE) Loss, while the regression branch incorporates the Distribution Focal Loss [22]. This loss function effectively addresses issues such as class imbalance and difficulty classifying challenging samples.

In terms of positive and negative sample assignment strategies, YOLOv8 departs from the static allocation strategy used in YOLOv5. Instead, it adopts the Task-Aligned Assigner from the TOOD [23] algorithm. This assigner selects positive samples based on the weighted scores of classification and regression tasks.

We chose the YOLOv8 algorithm as our baseline method for improvement due to its high accuracy, fast execution speed, and strong generalizability. Based on the analysis of the non-linear characteristics of the IP102 agricultural dataset and the distribution of the true bounding boxes of the targets, we have proposed improvements to the CBS module. These enhancements allow for better feature extraction and address the issue of gradient vanishing. Furthermore, we have introduced the DCF module to retain more low-level features of the targets. In the heatmap comparison presented in Section 4.2, we have compared our improved DCF layer with the C2F layer in the Yolov8 algorithm. Our algorithm exhibits superior feature extraction capabilities and achieves more accurate target detection on the IP102 dataset.

3. Materials and Methods

3.1. IP102 Dataset

The IP102 dataset was proposed by the Institute of Automation, Chinese Academy of Sciences. The dataset aims to provide a benchmark dataset for tasks such as image classification, object detection, and object segmentation. It consists of 102 categories, covering crops, fruits, vegetables, and other agricultural-related crops. The dataset has been widely used in the fields of object classification and detection.

3.1.1. Centralized Distribution

The IP102 dataset, being an agricultural dataset, consists mostly of images with a small number of pest categories and a relatively large proportion of the image occupied by pests, as depicted in Figure 3.

Figure 3. Distribution map of most pest targets in IP102 dataset.

Figure 4 presents the distribution of true bounding boxes for targets in the IP102 dataset. The x-axis and y-axis represent the x and y coordinates, while the height and width of the targets are depicted along the y-axis. The concentration of targets is primarily observed around the center of the images. The color intensity reflects the correlation between different targets. By combining Figures 3 and 4, it can be concluded that the targets in the images tend to be centrally located and exhibit a proportional relationship

between their width and height. Additionally, low-level features such as texture and shape demonstrate a significant advantage in the IP102 dataset.

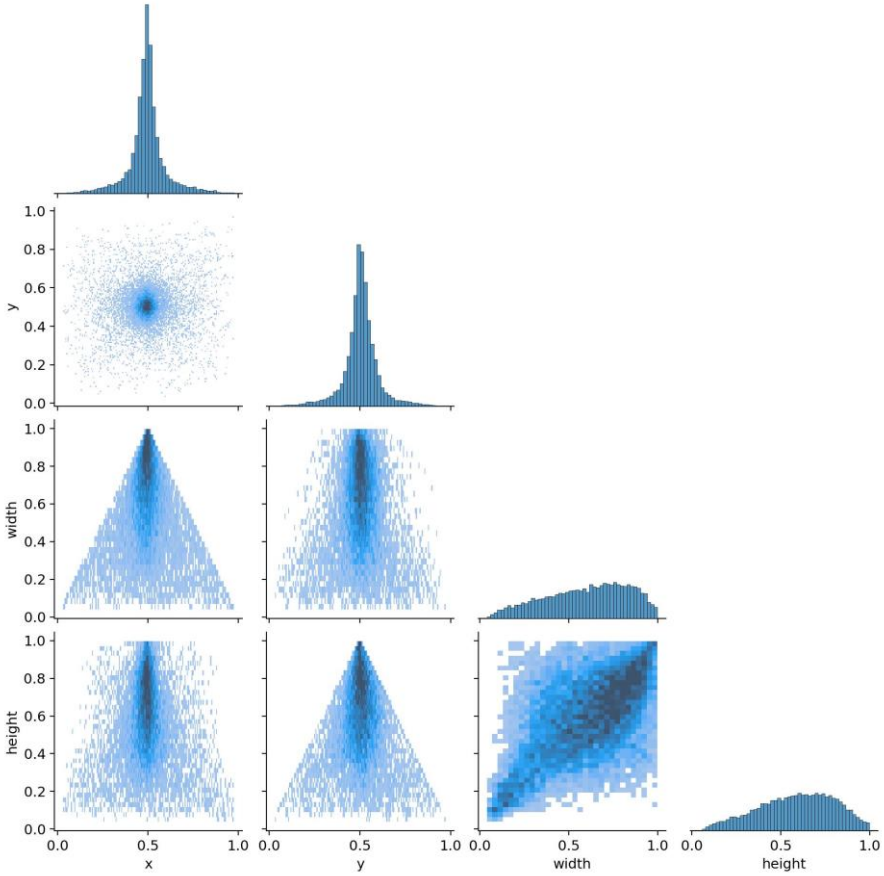


Figure 4. Target True Border Statistical Map.

3.1.2. Non-Linear Features

We selected a subset of the training dataset to obtain the histogram of pixel values, as shown in Figure 5. The histogram exhibits multiple sharp peaks in different intensity regions, with varying frequencies and irregular and asymmetric shapes. Therefore, the IP102 agricultural dataset exhibits nonlinear characteristics.

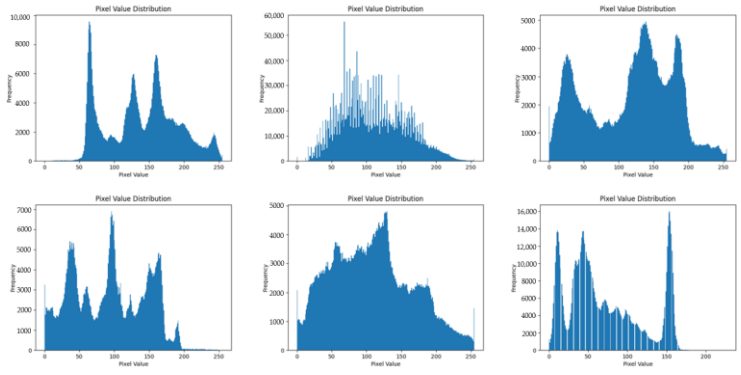


Figure 5. IP102 dataset pixel distribution histogram.

The Yolov8 algorithm model is relatively deep, and during the training of the IP102 dataset, when the number of training iterations reaches 50, the presence of non-linear features causes the gradients to decrease during the backpropagation process. As a result,

the model struggles to converge and fails to learn meaningful feature representations. The training process of the original algorithm is shown in Figure 6.

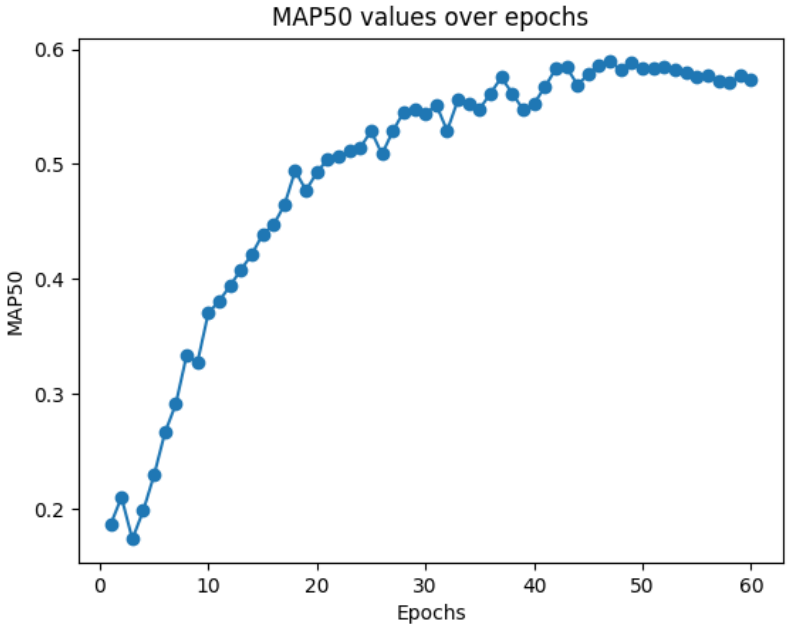


Figure 6. MAP50 training results with SiLU activation function.

3.2. Improved Model

In this paper, we propose improvements to the Yolov8 algorithm for the IP102 agricultural pest dataset, introducing the DCF module and the CBM module. The DCF module aggregates low-level features, enabling better feature extraction capabilities on the IP102 dataset. While the SiLU activation [24] function of the Yolov8 algorithm performs well on other datasets, it suffers from the gradient vanishing problem on the IP102 dataset. By replacing it with the Mish activation [25] function, our proposed model effectively addresses the gradient vanishing issue in deeper models. The improved algorithm model architecture is depicted in Figure 7. The DCF structure diagram is shown in Figure 9.

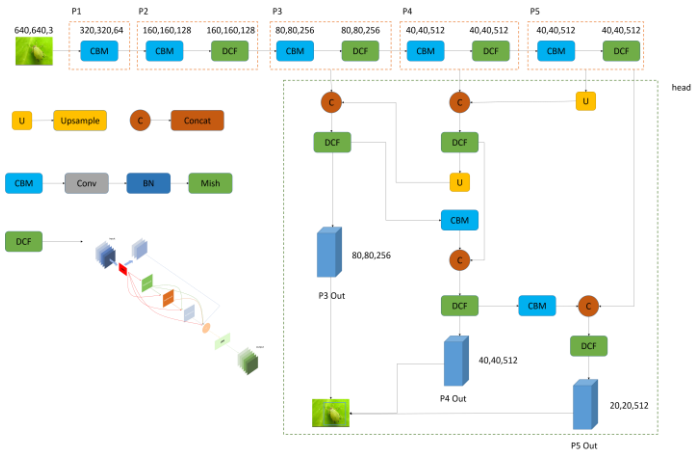
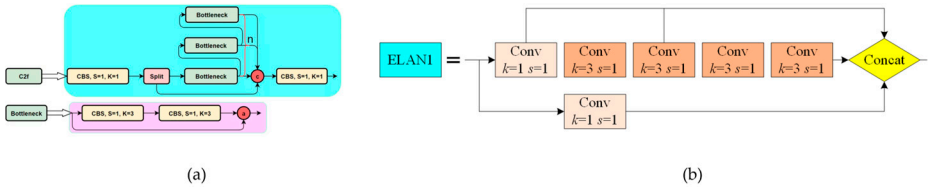


Figure 7. Improved Model Diagram.

3.2.1. Low-Level Feature Extraction Module DCF

The C2F module in Yolov8 draws inspiration from both the C3 module and the ELAN approach. It aims to achieve a lightweight design while capturing richer gradient flow information. Please refer to Figure 8 for a visual representation of the C2F module.



(a) (b)

Figure 8. (**a**) C2F structure diagram. (**b**) ELAN structure diagram.

As the network model reaches a certain depth, the accuracy gain diminishes when continuing to stack convolutional blocks, and the convergence deteriorates. The ELAN module solves this problem by only increasing the longest gradient path in residual blocks. By analyzing the shortest and longest gradient paths in each layer, the ELAN module controls the gradient paths to learn more features when the network becomes excessively deep.

The outputs of different bottlenecks in Yolov8 have distinct receptive fields and resolutions, which is advantageous for capturing multi-scale features and capturing details and semantic information at different levels. The C2F module demonstrates high accuracy in detection tasks involving multiple objects.

When there are fewer and larger objects in the detection images, low-level features are often more beneficial for model training. Prominent edge, texture, and shape features can be easily captured by low-level features, leading to accurate detection. High-level features contain more object composition and contextual information. However, in cases with fewer detection targets, high-level features do not exhibit significant advantages. Therefore, we propose an improvement to the C2F module based on the DenseBlock concept [26], called DCF, which retains its rich gradient flow information while emphasizing the low-level feature information of the targets. The DCF module is illustrated in Figure 9.

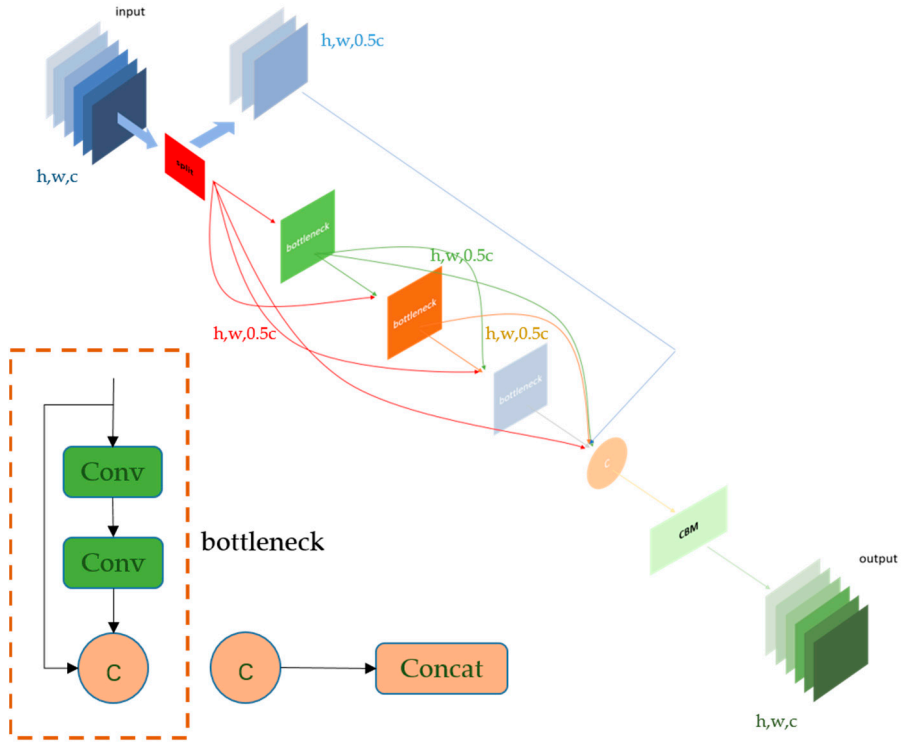


Figure 9. DCF Structure Diagram.

The DCF module allows each bottleneck to directly access the outputs of all previous layers, enabling the propagation of low-level features to subsequent bottlenecks. DCF can fully leverage previous low-level features and internally transmit them, so as the network model deepens, each bottleneck can obtain feature information from preceding layers,

enhancing the representation ability of low-level features. The "Split" operation transforms channel-wise features into spatial-wise features, thus better capturing low-level features such as edges and textures.

By employing cascaded connections, the low-level features extracted by earlier modules can be passed on to subsequent modules. The subsequent modules can consider both low-level and high-level features simultaneously, enabling a more comprehensive feature representation. Moreover, within the cascaded connections, the transmission and interaction of information help address the issue of information disappearance within the model, thereby improving the model's performance. With the replacement of DCF, the proposed algorithm in this paper achieves the best smoothness and the highest average detection accuracy.

3.2.2. Mish Activation Function

The convolutional module structure in YOLOv8 consists of convolutional blocks, normalization blocks, and the SiLU activation function.

The SiLU activation function approaches a linear function for input values that are either small or large. This allows the model to learn linear relationships more quickly, leading to faster convergence during the initial stages of training. In the case of using a pre-trained model, the YOLOv8 algorithm can achieve the highest detection accuracy with fewer epochs.

When training with the IP102 agricultural dataset, the training results for MAP50 are depicted in Figure 6. During the last 10 epochs of training, the MAP of the Yolov8 algorithm starts to decline and approaches saturation. Based on the nonlinear characteristics observed in the IP102 dataset and the poor performance of the SiLU function in handling gradient saturation, we replaced the CBS with CBM in Yolov8. Additionally, for the IP102 dataset, we employed the Mish activation function, whose function graph is shown in Figure 10.

$$f(x) = x * \frac{\left(e^{\ln(e^x+1)} - e^{(\ln(e^x+1)-1)}\right)}{\left(e^{\ln(e^x+1)} - e^{(\ln(e^x+1)*-1)}\right)} \quad (1)$$

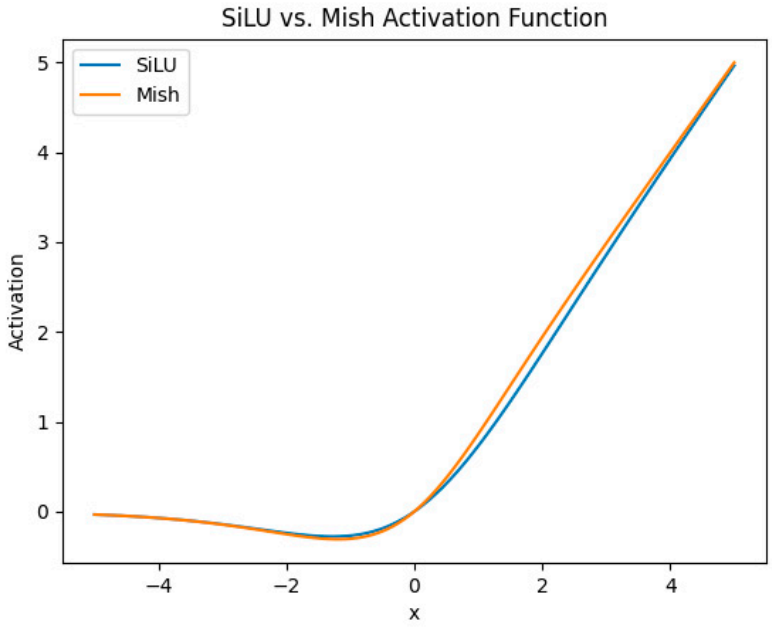


Figure 10. Curve comparison between SiLU activation function and Mish activation function.

The Mish activation function exhibits good smoothness and possesses differentiable continuity across the entire range of inputs. It effectively avoids the issues of gradients vanishing and exploding during the training process. Although the SiLU and Mish ac-

tivation functions have similar shapes, Mish has a higher curvature near zero in deeper models, which helps to avoid the gradient vanishing problem. Compared with SiLU, Mish can better capture the complex non-linear relationships present in the input data, which aligns well with the characteristics of the IP102 agricultural dataset. By better handling non-linear features, the Mish activation function enables Yolov8 to learn intricate features in deep networks, thereby improving the model's accuracy and generalization capabilities. The training results are shown in Figure 11.

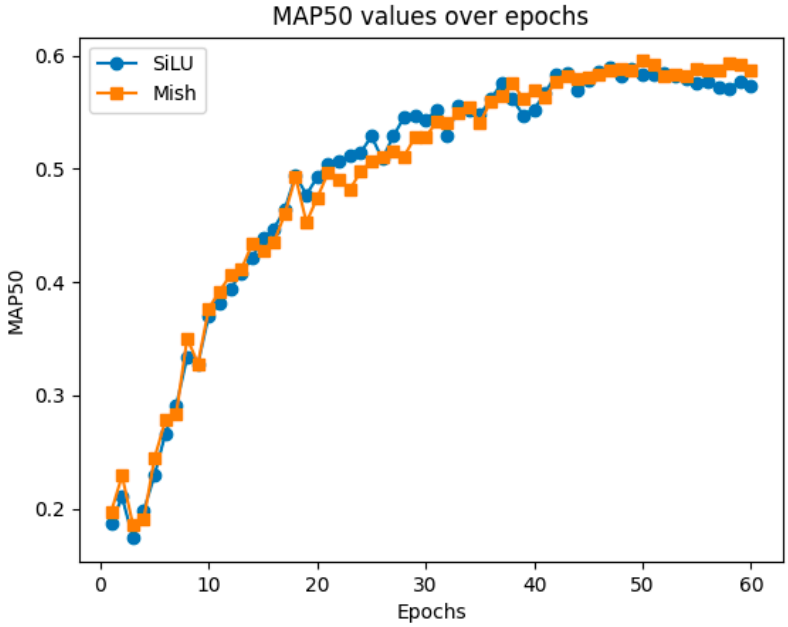


Figure 11. The training result graph of SiLU activation function and Mish activation function.

During the training process, the Mish activation function outperforms SiLU in terms of effectiveness. Mish training exhibits smoother performance in MAP50, enhancing the algorithm's detection results by better handling non-linear features. Please refer to Section 4.2 for the details of the ablation study.

4. Results

In this paper, we conducted experiments using the IP102 agricultural dataset. The IP102 dataset consists of image samples from 102 object categories, making it highly versatile. With a total of over 75,000 images, this dataset provides abundant samples for training and evaluation purposes. Additionally, the dataset encompasses multiple viewpoints, which effectively enhances the model's generalization and robustness.

The model evaluation metrics are as follows:

Precision: It measures the proportion of correctly detected targets among the ones detected by the model. A higher precision indicates more accurate detection by the model.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})} \tag{2}$$

In the equation TP represents the number of correctly detected targets. FP represents the number of falsely detected targets.

Recall measures the ability of a model to correctly detect targets. A higher recall indicates that the model can better capture the targets.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \tag{3}$$

In the equation FN is the number of targets that were not detected correctly.

AP: A comprehensive evaluation of the model's performance at different recall levels by calculating the average precision. It provides an overall assessment of the model's performance at various threshold values.

MAP50: The Average Precision (AP) at an Intersection over Union (IOU) threshold of 0.5 is calculated to measure the model's performance in predicting object locations accurately. It quantifies how well the model performs in terms of precision when the predicted bounding boxes have an IOU of at least 0.5 with the ground truth boxes.

MAP50-95: The Average Precision (AP) is calculated by considering the IOU values ranging from 0.5 to 0.95. This comprehensive evaluation takes into account the performance of the object detection model at different IOU thresholds and provides a more comprehensive assessment of its accuracy. It measures the precision of the model's predictions across a range of IOU thresholds, reflecting its ability to accurately localize objects under varying levels of overlap with the ground truth bounding boxes.

4.1. Model Training Results

The YOLOv8 algorithm, when using pre-trained weights, achieves optimal accuracy within 60 epochs. In this study, the proposed improved algorithm is trained for 60 epochs to perform ablation experiments and compared with the YOLOv8 algorithm. The training results of the proposed algorithm are shown in Figure 12.

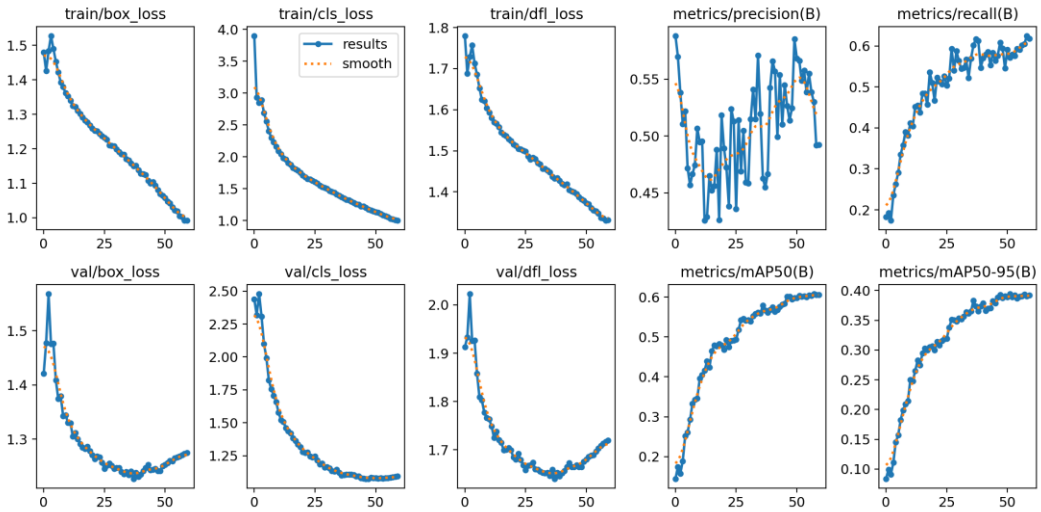
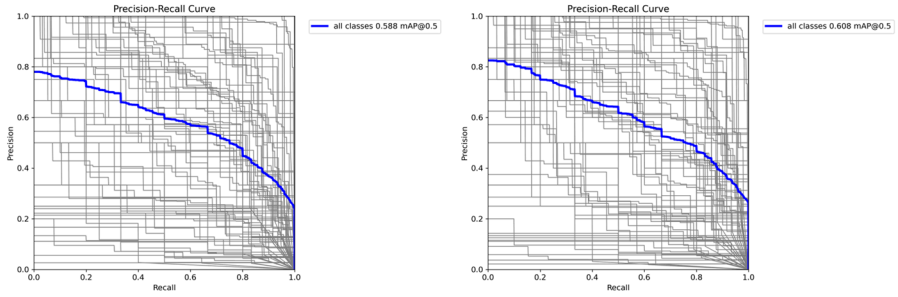


Figure 12. Model training result.

Figure 12 illustrates the variation curves of the loss on the training and validation sets, as well as the PR curves and MAP50 and MAP50-95 scores during the training of our proposed model.

Figure 13a depicts the PR curve of the YOLOv8 algorithm, while Figure 13b illustrates the PR curve of the proposed algorithm. The PR curve of the proposed algorithm encompasses the PR curve of the YOLOv8 algorithm, demonstrating that the proposed model surpasses the YOLOv8 model. Moreover, in terms of the area under the PR curve, specifically the MAP50 metric, the proposed algorithm exhibits a 2% improvement over the YOLOv8 algorithm.



(a) (b)

Figure 13. Model training PR curve. (**a**) is the PR curve of the YOLOV8 algorithm, (**b**) is the PR curve of the algorithm in this paper.

4.2. Ablation Experiment

Table 1 shows the training metrics of the Yolov8 algorithm on the IP102 dataset. With the use of a pre-trained model and the proposed improvements, including replacing CBS with CBM and adding the DCF module, our approach achieves better results even with fewer training epochs. The performance indicators clearly demonstrate the effectiveness of the proposed enhancements compared with the baseline Yolov8 algorithm.

Table 1. Ablation experiment.

Algorithm	MAP50	MAP50-95	P	R
Yolov8	58.8	39.4	51.7	56.7
CBM	59.5	39.3	53.6	58.6
CBM + DCF	60.8	39.4	53	60.4

During the training process, after replacing the CBS module with CBM and adding the DCF module, the training progress is shown in Figure 14. The SiLU activation function exhibits limited accuracy improvement when the model depth reaches its critical point, leading to the issue of gradient vanishing. By replacing it with the Mish activation function, the model can better handle the nonlinear features in the dataset and extract target features more effectively, resolving the gradient descent problem when the algorithm achieves optimal performance. Furthermore, replacing the C2F module with the DCF module further enhances the detection accuracy of the model trained with the Mish activation function. The DCF module aggregates low-level features, which contain rich and detailed information, to help the model better understand the data. Low-level features focus more on the local details of the image, making them more robust in handling complex scenes and large-scale images. In cases of model overfitting, an excessive reliance on high-level features can lead to increased model complexity. By using the optimized DCF module to extract low-level features, we can limit model complexity and reduce feature redundancy, thereby mitigating the overfitting phenomenon.

The improved DCF module in this paper focuses on low-level features and is suitable for the training dataset used in this study. A comparison between the heatmaps of the C2F module in Yolov8 and the DCF module proposed in this paper is shown in Figure 15. The heatmap is generated by utilizing the C2F layer from the original model and the DCF layer from our proposed algorithm. The effective extraction of good low-level features allows for better recognition of highly repetitive details in the image, resulting in superior texture information within the heatmap. Each pixel in the heatmap corresponds to the target or confidence score at the respective position. The more accurate the extraction of low-level features, the higher the scores assigned to texture and other information, leading to brighter and more prominent regions in the heatmap. The proposed algorithm achieves a 2% improvement in terms of Mean Average Precision (MAP50) compared with Yolov8.

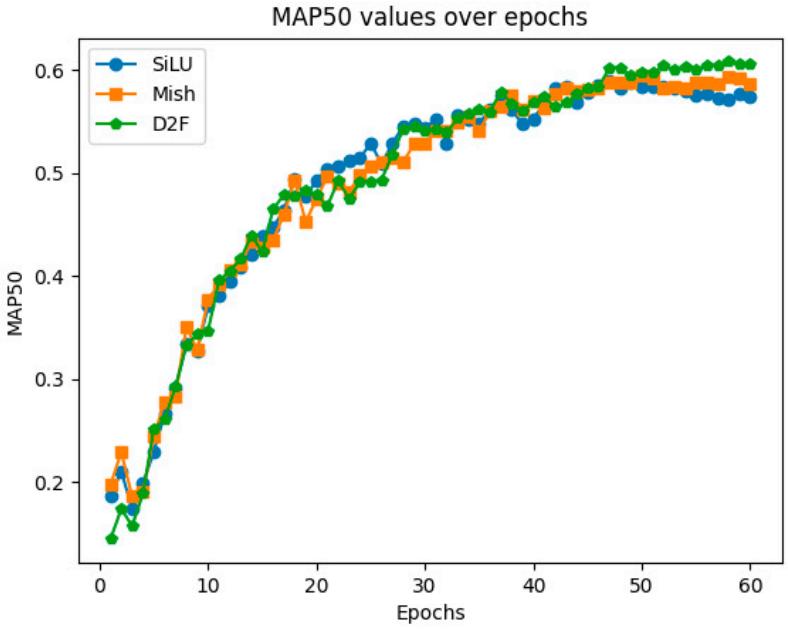


Figure 14. IP102 data set training result graph (the circle is the SiLU function training curve, the square is the Mish function training curve, and the five-pointed star is the DCF training curve).

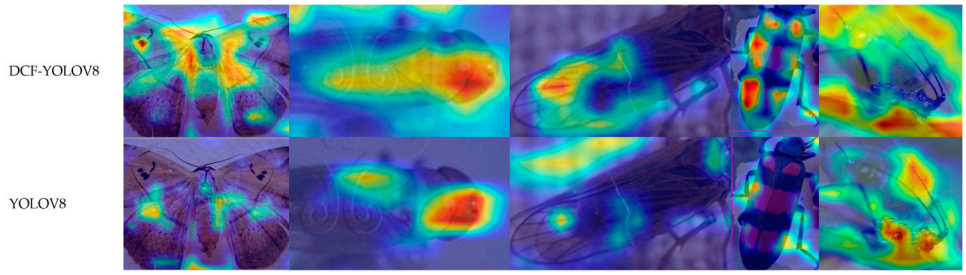


Figure 15. Experimental comparison heat map.

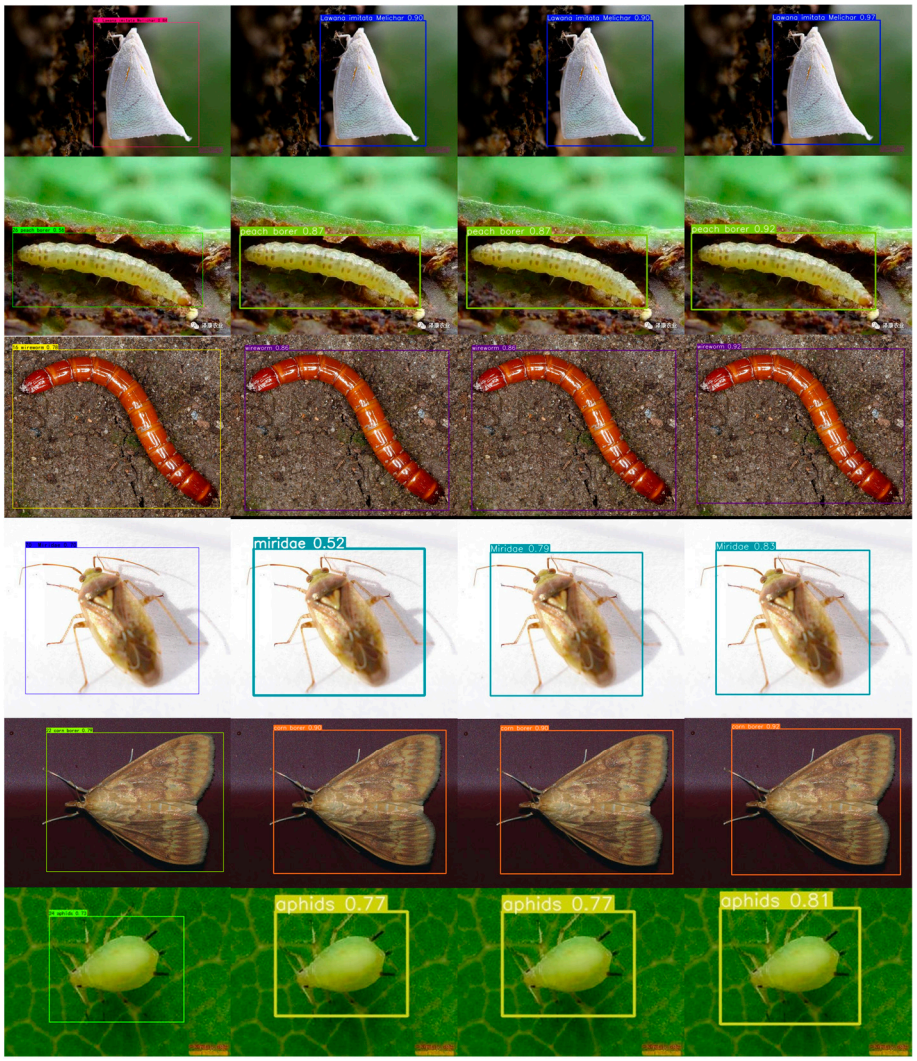
4.3. Comparative Experiment

In our comparative experiments, we selected FPN, Yolox, Dynamic R-CNN, SSD300, RefineDet, and ExquisiteNet as benchmark models. The experimental results are presented in Table 2. Our proposed algorithm demonstrates the best performance in terms of MAP50 and MAP50-95. On the IP102 dataset, our model exhibits higher accuracy and generalization capability compared with the other models.

Table 2. Comparative Experiment (* Represents coco data set training format indicators, backbone is Resnet50 [27]).

Algorithm	MAP50	MAP50-95	Params (M)	FPS
Yolov8	58.8	39.4	25.8	130
* DRCNN [28]	50.7	29.4	41.8	154
YoloX [29]	52.1	31.1	9.0	376
* CenterNet [30]	40.2	24.3	32.3	156
* SSD300 [31]	47.2	21.5	37.2	320
* FPN [32]	45.3	24.9	45	156
ExquistiteNet [13]	52.32	/	0.98	1933
Yolov5 [14]	56.2	34.1	7.2	238
* TOOD [23]	43.9	26.5	32.2	155
Ours	60.8	39.4	25.8	125

We conducted comparative experiments on the IP102 dataset, comparing our proposed improved algorithm with the Yolov8, YoloX, and Yolov5 algorithms. The results of the comparative experiments are shown in Figure 16. We specifically selected scenarios where the target background was similar, the target background was blurred, and the target had inconsistent lighting conditions for comparison. In different lighting conditions, the edges and textures of the target are difficult to capture, and low-level features are better at extracting these fine details. For complex scenarios with similar backgrounds or blurriness, our improved algorithm preserves the rich gradient flow information of the C2F model while aggregating low-level features. By fusing the detailed information from low-level features and the global semantic information from high-level features, the model gains a better understanding of the image. Additionally, it adjusts the weights based on different features to suppress noise and redundancy in complex backgrounds. Our proposed improved algorithm consistently achieved the highest score in all these scenarios.



(a) YoloX (b) Yolov5 (c) Yolov8 (d) Ours

Figure 16. Comparison chart of target detection results.

5. Discussion

This paper analyzes the characteristics of the IP102 agricultural pest dataset and proposes the DCF low-level feature extraction module and the CBM module. During training, the issue of gradient vanishing is effectively addressed, leading to more accurate extraction of features such as texture and edges related to agricultural pests.

Compared with YOLOv5, FPN, and YOLOX algorithms, the proposed DCF module in this paper better handles the texture, edges, and other information related to agricultural pests by transmitting low-level feature information. In Figure 15, the heatmap demonstrates that the proposed algorithm achieves higher scores at positions corresponding to texture edges. The Mish activation function exhibits advantages over SiLU in handling non-linear features. With a deep network architecture and the fusion of low-level and high-level features, the proposed model better understands agricultural pests, as depicted in Figure 16, where it outperforms other algorithms in complex background environments, achieving the best detection scores and the highest MAP50 accuracy. However, the IP102 dataset also contains scenarios with dense small-scale pests, where the proposed algorithm performs relatively poorly. Compared with algorithms such as AM-ResNet and ExquisiteNet, the proposed model has a larger parameter count and also faces limitations in terms of inference speed. The extraction of features from small objects and the development of lightweight structures are the directions for our future work.

6. Conclusions

The IP102 dataset, with its diverse and extensive collection of agricultural pests and diseases, is an ideal choice as the primary dataset for evaluating the proposed algorithm in this paper. The Yolov8 algorithm, being the current state-of-the-art algorithm with excellent performance, serves as the baseline for our improvements. Firstly, we conducted an analysis of the IP102 dataset and identified relevant features along with the limitations of the Yolov8 algorithm when trained on agricultural pest and disease datasets. Secondly, we proposed the CBM module to effectively handle the dataset's non-linear features and address the issue of gradient vanishing during model training. Thirdly, the introduction of the DCF module provided the algorithm with an advantage in extracting low-level features, leading to better representation of pest and disease textures, edges, and other characteristics. The fusion of low-level and high-level features enhanced the model's robustness in complex environments, resulting in significantly improved detection scores across the board.

Author Contributions: Methodology, G.D., L.Z. and D.L.; Dataset preparation, G.D. and C.L.; Experiments, G.D., L.Z. and D.L.; Original draft, L.Z. and C.L.; Review and editing, C.L. and G.D.; Visualization, G.D., L.Z. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the NSFC, grant number 61806024; Jilin Province Science and Technology Development Plan Key Research and Development Project, grant number 20210204050YY; Wuxi University Research Start-up Fund for Introduced Talents, grant numbers 2023r004, 2023r006.

Data Availability Statement: All the data mentioned in the paper are available through the corresponding author.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding the publication of this paper.

References

1. Ahmed, H.F.A.; Seleiman, M.F.; Mohamed, I.A.A.; Taha, R.S.; Wasonga, D.O.; Battaglia, M.L. Activity of Essential Oils and Plant Extracts as Biofungicides for Suppression of Soil-Borne Fungi Associated with Root Rot and Wilt of Marigold (*Calendula officinalis* L.). *Horticulturae* **2023**, *9*, 222. [CrossRef]
2. Ahmed, H.F.A.; Elnaggar, S.; Abdel-Wahed, G.A.; Taha, R.S.; Ahmad, A.; Al-Selwey, W.A.; Ahmed, H.M.H.; Khan, N.; Seleiman, M.F. Induction of Systemic Resistance in Hibiscus sabdariffa Linn. to Control Root Rot and Wilt Diseases Using Biotic and Abiotic Inducers. *Biology* **2023**, *12*, 789. [CrossRef] [PubMed]
3. Chaudhary, A.; Kolhe, S.; Kamal, R. An improved random forest classifier for multi-class classification. *Inf. Process. Agric.* **2016**, *3*, 215–222. [CrossRef]
4. Singh, A.K.; Sreenivasu, S.V.N.; Mahalaxmi, U.; Sharma, H.; Patil, D.D.; Asenso, E. Hybrid feature-based disease detection in plant leaf using convolutional neural network, bayesian optimized SVM, and random forest classifier. *J. Food Qual.* **2022**, *2022*, 2845320. [CrossRef]
5. Panchal, P.; Raman, V.C.; Mantri, S. Plant diseases detection and classification using machine learning models. In Proceedings of the 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 20–21 December 2019; Volume 4, pp. 1–6.

6. Meenakshi, M.; Naresh, R. Soil health analysis and fertilizer prediction for crop image identification by Inception-V3 and random forest. *Remote Sens. Appl. Soc. Environ.* **2022**, *28*, 100846. [CrossRef]
7. Ren, F.; Liu, W.; Wu, G. Feature reuse residual networks for insect pest recognition. *IEEE Access* **2019**, *7*, 122758–122768. [CrossRef]
8. Wu, X.; Zhan, C.; Lai, Y.K.; Cheng, M.M.; Yang, J. Ip102: A large-scale benchmark dataset for insect pest recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8787–8796.
9. Nanni, L.; Maguolo, G.; Pancino, F. Insect pest image detection and recognition based on bio-inspired methods. *Ecol. Inform.* **2020**, *57*, 101089. [CrossRef]
10. Kasinathan, T.; Singaraju, D.; Uyyala, S.R. Insect classification and detection in field crops using modern machine learning techniques. *Inf. Process. Agric.* **2021**, *8*, 446–457. [CrossRef]
11. Feng, Y.; Liu, Y.; Zhang, X.; Li, X. TIR: A Two-Stage Insect Recognition Method for Convolutional Neural Network. In Proceedings of the Pattern Recognition and Computer Vision: 5th Chinese Conference, PRCV 2022, Shenzhen, China, 4–7 November 2022; Proceedings, Part II. Springer Nature: Cham, Switzerland, 2022; pp. 668–680.
12. Zhang, L.; Du, J.; Dong, S.; Wang, F.; Xie, C.; Wang, R. AM-ResNet: Low-energy-consumption addition-multiplication hybrid ResNet for pest recognition. *Comput. Electron. Agric.* **2022**, *202*, 107357. [CrossRef]
13. Zhou, S.Y.; Su, C.Y. Efficient convolutional neural network for pest recognition-ExquisiteNet. In Proceedings of the 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 23–25 October 2020; pp. 216–219.
14. Glenn, J.; Alex, S.; Ayush, C.; Jirka, B. *Ultralytics/Yolov5: V6.0—YOLOv5n "Nano" Models, Roboflow Integration, TensorFlow Export, OpenCV DNN Support, Version 6.0*; Zenodo: Honolulu, HI, USA, 2021.
15. Lyu, S.; Ke, Z.; Li, Z.; Xie, J.; Zhou, X.; Liu, Y. Accurate Detection Algorithm of Citrus Psyllid Using the YOLOv5s-BC Model. *Agronomy* **2023**, *13*, 896. [CrossRef]
16. Feng, J.; Yu, C.; Shi, X.; Zheng, Z.; Yang, L.; Hu, Y. Research on Winter Jujube Object Detection Based on Optimized Yolov5s. *Agronomy* **2023**, *13*, 810. [CrossRef]
17. Lou, L.; Liu, J.; Yang, Z.; Zhou, X.; Yin, Z. Agricultural Pest Detection based on Improved Yolov5. In Proceedings of the 2022 6th International Conference on Computer Science and Artificial Intelligence, Beijing, China, 9–11 December 2022; pp. 7–12.
18. Doan, T.N. An Efficient System for Real-time Mobile Smart Device-based Insect Detection. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*. [CrossRef]
19. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
20. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.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, 18–22 June 2023; pp. 7464–7475.
21. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
22. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
23. Feng, C.; Zhong, Y.; Gao, Y.; Scott, M.R.; Huang, W. Tood: Task-aligned one-stage object detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE Computer Society, Montreal, BC, Canada, 11–17 October 2021; pp. 3490–3499.
24. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941.
25. Misra, D. Mish: A self regularized non-monotonic activation function. *arXiv* **2019**, arXiv:1908.08681.
26. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 4700–4708.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
28. Zhang, H.; Chang, H.; Ma, B.; Wang, N.; Chen, X. Dynamic R-CNN: Towards high quality object detection via dynamic training. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part XV 16. Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 260–275.
29. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
30. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6569–6578.
31. 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; Proceedings, Part I 14. Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
32. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.