Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.33427

Automatic Detection of Plant Leaf Diseases Based on Improved Yolov8 Algorithm

Qinglong Zhang[1], Zhongliang Kan[1*]

[1]*School of Computer Science and Big Data, Heilongjiang University, Harbin, China*
**Corresponding Author. Email: 1992007@hlju.edu.cn*

Abstract. In this paper, we introduce two-way weighted feature pyramid and convolutional block attention mechanism based on the single-stage target detection framework to enhance multi-scale feature fusion and highlight the lesion-related response. The experiment adopted the configuration of batch size of 8 and a total of 50 rounds of training, and carried out training and evaluation on the data division containing 1136 training images and 211 validation images. The validation set contained 769 labeled instances and covered more than ten types of pests and diseases. During the training process, the loss of bounding box, classification and distribution focus decreased overall, while the accuracy rate, recall rate and mAP index of validation set increased or fluctuated with the training progress. Ablation experiments show that adding attention module and feature pyramid structure to the baseline model can bring accuracy benefit, and the combination of the two can achieve better detection performance. At the same time, the number of parameters and calculation overhead increase, and the inference speed decrease slightly. The results show that the proposed improvement is beneficial to disease localization and identification in complex background and small target scenarios, and can provide a reference for visual monitoring in smart agriculture.

Keywords: Plant leaf disease detection, two-way weighted feature pyramid, CBAM, YOLOv8

1. Introduction

Global agricultural production has been threatened by leaf diseases caused by pathogens such as fungi and bacteria for a long time. The lesions are often small in area at the early stage, and are easily confused with shadows, insect bites or nutrient deficiency spots. Delayed control will significantly affect yield and quality [1]. The traditional method relies on field inspection and empirical judgment by agricultural technicians, which has limited coverage and high labor intensity, and it is difficult to unify standards among different observers [2]. With the popularization of digital cameras, mobile terminals and drones in the field, automatic disease analysis based on visible light images has become an important direction of smart agriculture; Among them, target detection can give the location and category of lesions in a single image at the same time, which is more conducive to describing the spatial distribution of diseases than simple whole image classification,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.33427

and provides a direct basis for variable application and early warning, so it has received wide attention [3].

Machine learning, especially supervised deep learning, establishes a complex nonlinear mapping from pixels to disease semantics by optimizing the loss function on large-scale labeled data, so that the system can automatically learn category-related representations such as leaf color, texture, edge, and lesion morphology without completely relying on artificially designed rule features [4]. Through layer-by-layer abstraction, convolutional neural networks can adapt to actual interference such as lighting changes, shooting angles and complex backgrounds to a certain extent, and have the ability to deploy on the edge or cloud, batch reasoning and continuous iteration, thus reducing monitoring costs and improving results consistency, providing a scalable technical path for large-scale and high-frequency screening of agricultural diseases; Its performance largely depends on the network's ability to adequately fuse multi-scale information and highlight lesion-related responses [5]. Under this background, the single-stage detection algorithm represented by YOLOv8 has shown the advantages of clear training process and fast reasoning speed in the open source community plant disease task. However, under the circumstances of large lesion scale differences and staggered background branches and leaves, there may still be problems such as insufficient utilization of multi-scale features and unemphasized key areas.

Therefore, this paper introduces BiFPN feature pyramid structure on the basis of YOLOv8 detection framework, and improves the collaborative expression of features with different resolutions through cross-scale bidirectional fusion and learnable weights; And the CBAM attention module is embedded to adaptively recalibrate the features in the channel and spatial dimensions to enhance the lesion area and suppress the irrelevant background. The two cooperate with the original detection heads to form an improved scheme for plant disease scenarios, striving to improve the detection accuracy and robustness while maintaining real-time performance, and providing a reference for disease intelligent diagnosis and engineering application.

2. Method and materials

2.1. Introduction to datasets

The leaf image dataset used in this paper was collected with digital imaging equipment under controlled conditions in the laboratory or in the natural scene in the field. After screening for clarity and usability, the lesion area was retained. The data processing process includes uniformly scaling the image to the input resolution of the model, organizing the bounding box and category number according to YOLO format, dividing the training set and the validation set according to scale, and supplemented by enhancement strategies such as random flip and scaling, mosaic stitching, and color and brightness disturbance to alleviate overfitting and cover imaging differences. In terms of category composition, the data cover dozens of fine-grained categories such as various crops and their health and typical diseases. The training set contains a total of 1136 images, and the validation set contains 211 images. When the two do not overlap with each other, the whole library totals about 1347 images, which are used for model learning and independent evaluation.

2.2. BiFPN

BiFPN is a weighted bidirectional pyramid network developed on the multi-scale feature fusion framework. Its core idea is to repeatedly transfer and fuse features with different resolutions in top-down and bottom-up directions, instead of relying only on one-way top-down paths [6]. Learnable

DOI: 10.54254/2755-2721/2026.33427

scalar weights are introduced in each layer of fusion, and the input features from different scales are normalized and weighted summed, so that the network can adaptively emphasize the scale contribution with greater information. At the same time, BiFPN enhances the interaction between shallow details and deep semantics while controlling computational overhead by deleting nodes with only a single input edge and adding additional cross-layer connections at the same resolution level [7]. Compared with the traditional feature pyramid network, this structure improves the fusion efficiency and expression ability while maintaining the multi-scale detection capability, and is suitable for visual tasks where the target scale changes significantly and needs to make fine use of middle-level features. The network structure of BiFPN is shown in Figure 1.

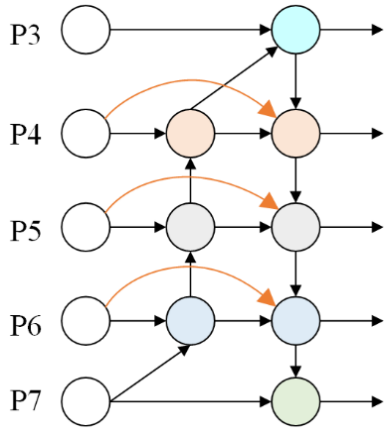


Figure 1. The network structure of BiFPN

2.3. CBAM attention module

CBAM is a lightweight convolutional attention mechanism that sequentially recalibrates feature maps along two dimensions, channel and space, to highlight task-related responses and suppress irrelevant backgrounds [8]. The branch of channel attention compresses the spatial dimension through global pooling, and models the importance weight of each channel through multi-layer perceptron, so that the network reinforces the response to discriminant channels at the semantic level. After the channel information is aggregated, the spatial attention branch estimates the weight of each spatial position on the feature map, and guides the network to pay attention to the region where the target may appear. Two sub-modules act on the same feature in series, the amount of parameters and the amount of computation are relatively controllable, and it is easy to embed different depths of convolutional backbone or neck network, thus improving the separability and positioning stability of features without significantly increasing the inference burden. The network structure of the CBAM attention module is shown in Figure 2.

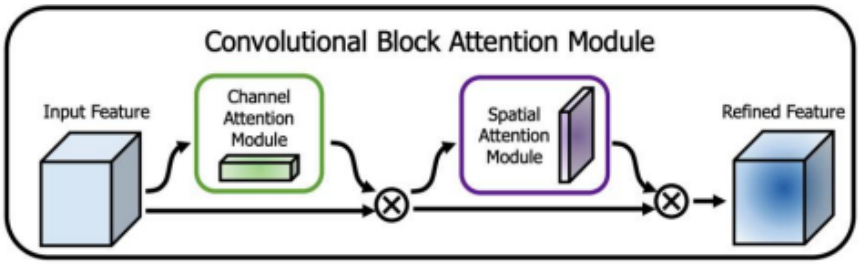


Figure 2. The network structure of the CBAM attention module

DOI: 10.54254/2755-2721/2026.33427

2.4. YOLOv8 algorithm

YOLOv8 belongs to the single-stage target detection paradigm. The image is input into the convolutional backbone network to extract multi-layer features from shallow to deep, and then multi-scale fusion is performed through the neck structure. Finally, the detection head predicts the category and position of the target in parallel on multiple feature maps [9]. In the structural design of this series, cross-stage partial connection modules are generally adopted to balance the accuracy and speed. The detection head usually adopts decoupling form to regress the bounding box and classification logits respectively, and uses distributed bounding box parameterization to better characterize the position and shape uncertainty of the frame. In the training stage, the classification loss, positioning loss and distribution-related loss are jointly optimized, and the generalization performance is improved with data enhancement and multi-scale training strategies. The whole process can be differentiated from end to end, and dense candidate results can be output by single forward reasoning, and the final detection frame can be obtained after non-maximum value suppression, taking into account both real-time performance and detection accuracy. The network structure of YOLOv8 is shown in Figure 3.

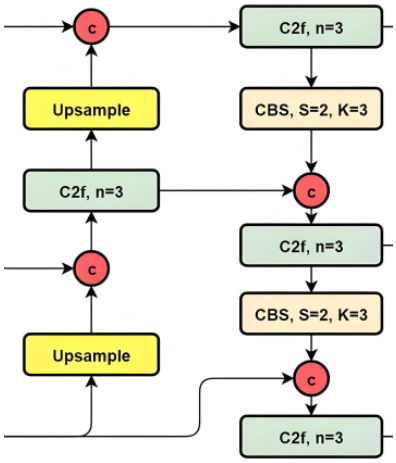


Figure 3. The network structure of YOLOv8

2.5. YOLOv8 algorithm based on BiFPN and CBAM optimization

The introduction of BiFPN above the YOLOv8 baseline aims to replace or enhance the original neck multi-scale fusion path, so that features from different depths of the backbone can obtain more adequate information interaction in bidirectional weighted fusion, thus improving the feature alignment problem between small targets and targets with larger scale span [10]. The CBAM module is arranged on the feature map after backbone output or BiFPN fusion, and the fused features are adaptively emphasized through channel and spatial attention, so that the lesion or target area can obtain higher response weight and weaken the complex background interference. The two are trained together with the original detection head and loss function. In the forward process, the features are first fused by pyramid and then refined by attention. In the back propagation, the gradient can update the fusion weight, attention parameters and detection head parameters at the same time. On the premise of maintaining the single-stage end-to-end detection process of YOLOv8, this combination improves the quality of multi-scale characterization and key region sensitivity from the two levels of structural fusion and feature recalibration. The network structure of the YOLOv8 algorithm based on BiFPN and CBAM optimization is shown in Figure 4.

DOI: 10.54254/2755-2721/2026.33427

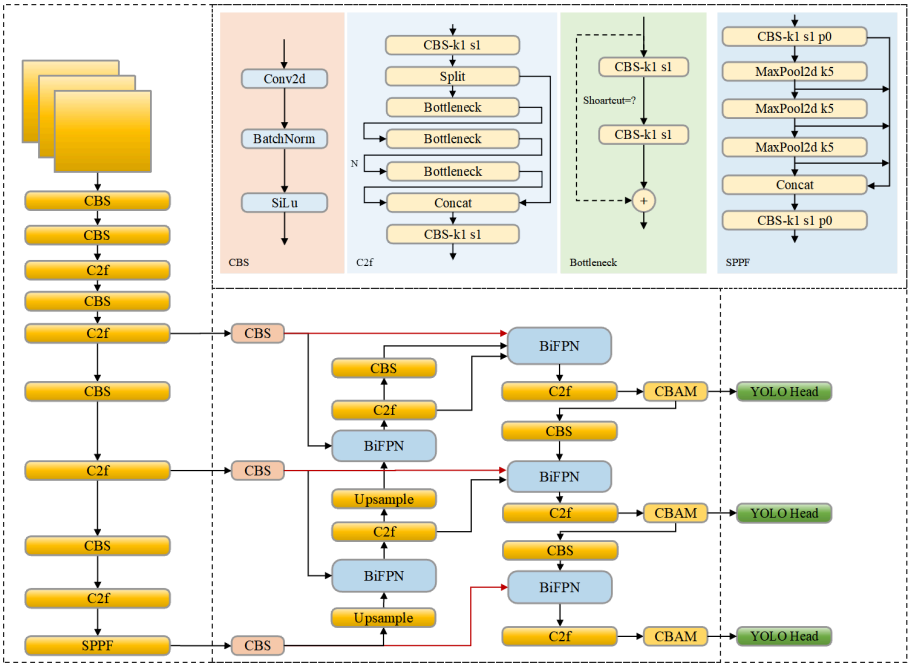
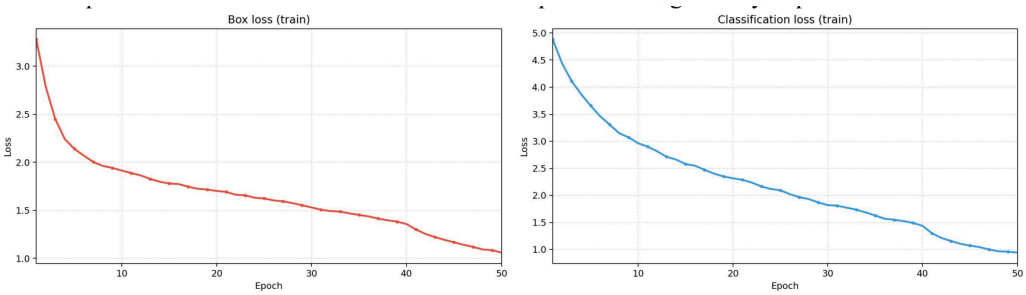


Figure 4. The network structure of the YOLOv8 algorithm based on BiFPN and CBAM optimization

3. Result

In the setting of experimental parameters, the training rounds are set to 50, the input image size is set to 800, and the training iterations of each epoch are 142 batches. When the batch size is 8, it is consistent with the size of about 1136 training samples per round. There are 7 batches in the verification stage, covering 211 verification images and 769 annotation instances. During the training process, the single card video memory occupies about 10.4 GB; The learning rate adopts an initial value of 0.01, which is decayed to about 0.0001 by cosine scheduling, with a momentum coefficient of 0.937 and a weight decay of 0.0005.

The results of various indicators of the training set are shown in Figure 5. During the training process, the three types of losses continue to decrease with epoch, and the accuracy, recall, mAP50 and mAP50-95 on the validation set generally show an upward or fluctuating upward trend, indicating that the model converges well on the plant disease detection task and the detection performance gradually improves.



DOI: 10.54254/2755-2721/2026.33427

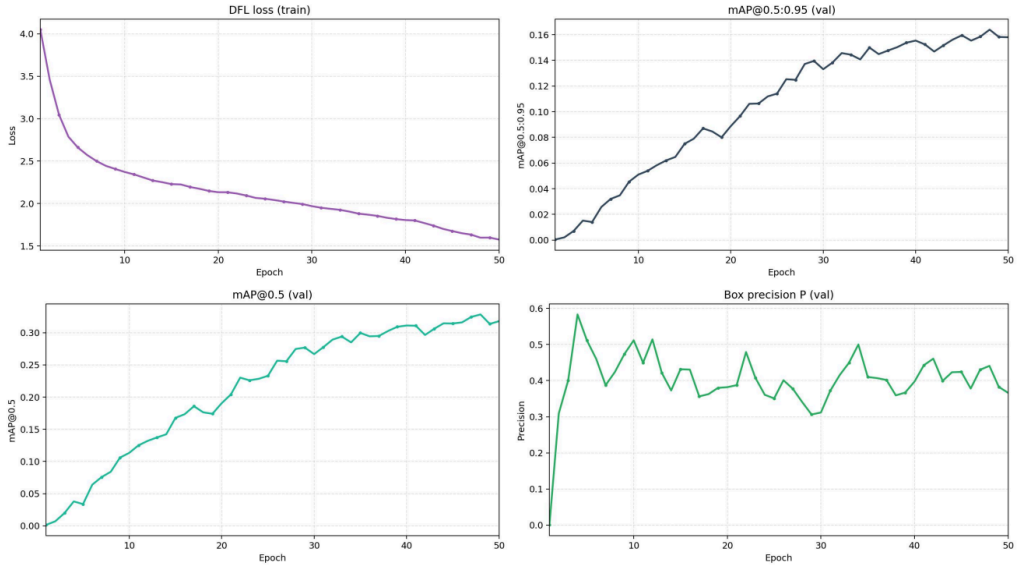


Figure 5. The results of various indicators of the training set

Output the results of the validation set as shown in Table 1.

Table 1. The results of the validation set

Class	Images	Instances	Box P	R	mAP50	mAP50-95
all	211	769	0.452	0.33	0.332	0.166
Aphids	211	20	0.539	0.05	0.0501	0.0247
Early Blight	211	96	0.455	0.312	0.319	0.151
Healthy Leaf-	211	40	0.387	0.35	0.311	0.171
Leaf Curl	211	14	0	0	0.00894	0.00396
Leafhoppers and Jassids	211	4	0.525	0.25	0.422	0.239
Molds	211	64	0.492	0.266	0.284	0.116
Mosaic Virus	211	98	0.65	0.571	0.584	0.286
Septoria	211	62	0.358	0.405	0.285	0.146
bactarial canker	211	30	0.411	0.3	0.283	0.124
bacterial spot	211	68	0.387	0.471	0.318	0.145
flea beetle	211	15	0.419	0.243	0.284	0.119
late_blight	211	114	0.705	0.605	0.686	0.345
leafminer	211	28	0.427	0.25	0.22	0.125
powedry_mildew	211	61	0.435	0.344	0.344	0.131
yellow curl virus	211	55	0.584	0.537	0.574	0.357

The results of the ablation experiments are shown in Table 2.

DOI: 10.54254/2755-2721/2026.33427

Table 2. The results of the ablation experiments

Method	P	R	mAP50	mAP50-95
YOLOv8	0.461	0.338	0.345	0.174
YOLOv8+ CBAM	0.474	0.346	0.358	0.181
YOLOv8+ BiFPN	0.481	0.352	0.366	0.185
YOLOv8s+ CBAM + BiFPN	0.496	0.361	0.379	0.193

The result of target detection of the validator of the YOLOv8s + CBAM + BiFPN algorithm is output, as shown in FIG. 6.

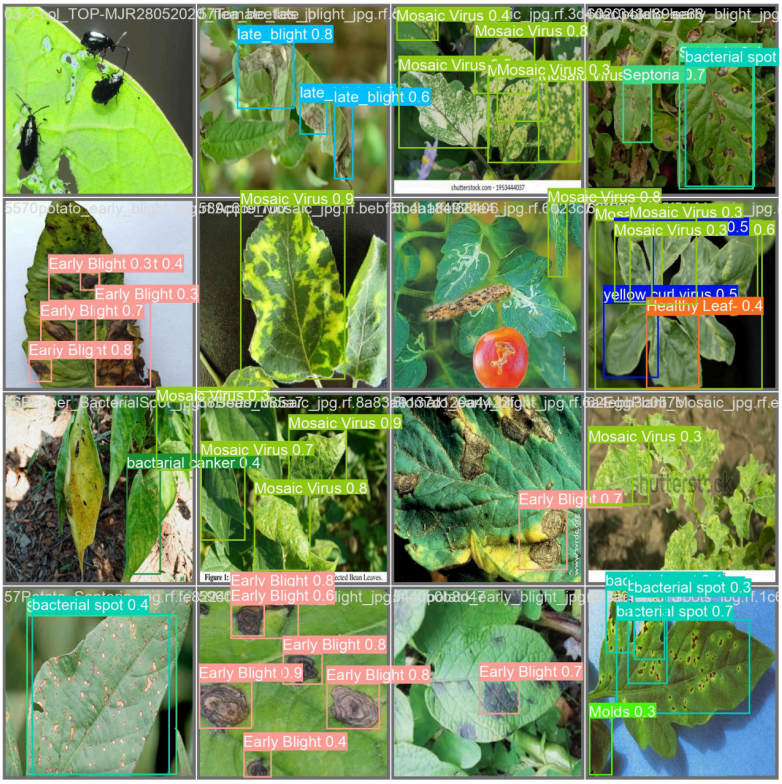


Figure 6. The result of target detection of the validator of the YOLOv8s + CBAM + BiFPN algorithm

The model on the validation set achieved mAP50 of the order of 0.35 and mAP50-95 of the order of 0.17 in the multi-category plant disease detection task of 211 images and 769 instances, and the overall accuracy and recall rate were at the middle level, indicating that it still has certain detection ability but still has room for improvement under the complex leaf background and scale differences. When a few categories have few samples or lesions are extremely difficult to distinguish from the background, the index is low, while categories with sufficient samples and obvious lesion characteristics can achieve higher accuracy and recall. Ablation experiments show that after introducing CBAM and BiFPN in sequence above the baseline of YOLOv8s, multi-scale fusion and channel spatial attention bring stable accuracy gain. When the two are combined, they reach the optimum on mAP and mAP50-95. At the same time, the amount of parameters and calculation increases, and the inference frame rate decreases slightly, which reflects the compromise between accuracy and efficiency.

DOI: 10.54254/2755-2721/2026.33427

4. Conclusion

In this study, the bidirectional feature pyramid and channel spatial attention are embedded in the neck of the detection network and the feature path, and the missed detection and false detection problems caused by scale disparity and background interference in the blade image are alleviated from the two levels of multi-scale fusion and feature recalibration. The experimental curves and quantitative indexes show that the model can converge stably on the set data partitioning, and obtain competitive detection accuracy on the validation set. The sample imbalance between categories still constrains the performance of some rare categories, which can be mitigated later by resampling, loss reweighting or introducing more fine-grained data augmentation. In addition, extrapolation verification on more field scenarios and different imaging devices has yet to be carried out to further verify the robustness and real-time performance of the model in real deployment environments. Generally speaking, this work achieves an acceptable compromise between accuracy and complexity, and provides reproducible experimental ideas and improvement directions for engineering application of plant disease target detection.

References

[1] Sharma, Sangeeta. "Apple leaf disease prediction using modified YOLOv8 algorithm." 2024 International Conference on Integrated Circuits, Communication, and Computing Systems (ICIC3S). Vol. 1. IEEE, 2024.
[2] Luo, Dehuan, et al. "Citrus diseases and pests detection model based on self-attention YOLOV8." IEEE access 11 (2023): 139872-139881.
[3] Chen, Ming, et al. "Tomato leaf disease detection method based on improved YOLOv8n." Scientific Reports 15.1 (2025): 25837.
[4] Zhang, Wei, et al. "Corn leaf disease spot image classification and recognition algorithm based on improved Yolov8." International Conference on Computer Application and Information Security (ICCAIS 2024). Vol. 13562. SPIE, 2025.
[5] Orchi, Houda, et al. "Real-time detection of crop leaf diseases using enhanced YOLOv8 algorithm." 2023 International Wireless Communications and Mobile Computing (IWCMC). IEEE, 2023.
[6] Chen, Chen, et al. "Research on soybean leaf disease recognition in natural environment based on improved Yolov8." Frontiers in Plant Science 16 (2025): 1523633.
[7] Uddin, Md Sazid, et al. "Cauli-Det: enhancing cauliflower disease detection with modified YOLOv8." Frontiers in Plant Science 15 (2024): 1373590.
[8] Zhou, Jinxuan, and Gang Nie. "Research and design of a crop leaf disease diagnosis system based on the YOLOV8 model." 2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT). IEEE, 2024.
[9] Preanto, Sabit Ahamed, et al. "A semantic segmentation approach on sweet orange leaf diseases detection utilizing YOLO." arXiv preprint arXiv: 2409.06671 (2024).
[10] Kaur, Ramanjot, et al. "YOLO-LeafNet: a robust deep learning framework for multispecies plant disease detection with data augmentation." Scientific Reports 15.1 (2025): 28513.