ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

Research on BoShao Recognition Based on Deep Learning

Xing Chunyu[1,a,*], Chen Jixiang[1], Liu Ling[1], Tan Jun[1], Yin Wei[1]

[1]*Department of Information Engineering, Bozhou Vocational and Technical College, Bozhou, China*
[a]*thuwsj@foxmail.com*
**Corresponding author*

Abstract: *To address the issue of BoShao (Paeonia lactiflora) being frequently substituted with inferior varieties in the traditional Chinese medicine market, this study proposes an enhanced deep learning-based recognition method by refining YOLOv8, aiming to overcome the inefficiency of manual detection and the challenges of small object recognition under complex backgrounds. The proposed approach integrates a dual channel-spatial attention mechanism (CBAM) and a pyramid-like localized patch network (PLPNet) to optimize YOLOv8's feature extraction capabilities: CBAM enhances perception of critical regions through channel-wise weighting and spatial focusing, while PLPNet improves fine-grained feature capture via multi-scale localized patch fusion. Experiments were conducted on a dataset comprising 1,659 field-collected BoShao images, augmented with preprocessing techniques such as random occlusion and Gaussian noise. Ablation experiments demonstrated that the YOLOv8+CBAM+PLPNet model achieved mAP@0.5 and mAP@0.95 scores of 98.66% and 60.70%, respectively, representing improvements of 2.28% and 4.81% over the baseline model. In comparative experiments, the proposed model outperformed state-of-the-art detectors including YOLOv11 and EfficientDet, achieving superior precision (98.69%) and recall (98.18%). The results confirm that the synergistic optimization of attention mechanisms and multi-scale feature fusion significantly enhances robustness in BoShao recognition under complex environmental conditions. This work provides an efficient automated solution for quality inspection of medicinal materials in practical agricultural applications.*

Keywords: *Attention mechanism, Multi-scale feature fusion, YOLOv8, Quality inspection of medicinal materials*

1. Introduction

In the traditional Chinese medicine market, BoZhao Shao (BoShao) flowers are frequently substituted with inferior varieties, compromising the quality of medicinal materials and undermining market credibility. Traditional identification methods relying on manual expertise and microscopic analysis are time-consuming and error-prone, failing to meet industry demands. While deep learning-based object detection models like YOLOv8 offer efficient solutions for BoShao recognition, their performance is hindered by challenges such as small object detection, blurred boundaries, and complex background interference[1].

To address these limitations, this study enhances YOLOv8 through two synergistic innovations. First, the Convolutional Block Attention Module (CBAM) is integrated to refine feature extraction by adaptively weighting critical regions of the image, enabling the model to prioritize small and boundary-ambiguous objects[2]. Second, a Pyramid-like Localized Patch Network (PLPNet) is introduced to strengthen multi-scale feature fusion, enhancing sensitivity to fine-grained details in complex environments[3]. By combining attention-driven localization with hierarchical feature learning, the proposed framework significantly improves both accuracy and robustness in BoShao recognition. This approach not only advances automated quality inspection in agricultural settings but also demonstrates the potential of hybrid architectures for challenging object detection tasks.

2. Research Methods

YOLOv8, as a powerful object detection model, has achieved remarkable results in many application scenarios. However, it still has limitations when dealing with some challenging tasks, especially in small object detection, blurred object boundaries, and complex background interference. To address these

ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

issues, this paper proposes the introduction of CBAM and PLPNet to improve the YOLOv8 backbone network, further enhancing its detection accuracy and robustness.

The traditional YOLOv8 backbone network is based on convolutional neural networks (CNNs), which extract feature maps from input images and perform object detection. However, Y To address YOLOv8's limitations in small object detection and blurred boundary handling (as mentioned in Section 1), this paper proposes integrating CBAM and PLPNet to refine the backbone network, which ultimately affects detection performance.

Mathematically, the object detection process in YOLOv8 can be represented by regression. Given the input image F_{input}, YOLOv8 generates the feature map F_{yolov8} through a convolutional network, and predicts the bounding box $\hat{b}$, class $\hat{c}$, and confidence $\hat{p}$:

$$F_{\text{yolov8}} = \text{CNN}(F_{\text{input}})$$

$$\hat{y} = \{\hat{c}, \hat{b}, \hat{p}\}$$

where $\hat{b} = (x, y, w, h)$ represents the predicted bounding box coordinates, $\hat{c}$ is the predicted class, and $\hat{p}$ is the confidence of the object.

The loss function in YOLOv8 consists of multiple components, mainly the bounding box regression loss, class loss, and confidence loss:

$$\mathcal{L} = \mathcal{L}_{obj} + \mathcal{L}_{noobj} + \mathcal{L}_{class} + \mathcal{L}_{bbox}$$

where $\mathcal{L}_{obj}$ is the loss for the presence of an object, $\mathcal{L}_{noobj}$ is the loss for the absence of an object, $\mathcal{L}_{class}$ is the class prediction loss, and $\mathcal{L}_{bbox}$ is the bounding box regression loss.

While YOLOv8 performs well in detecting large objects, the bounding box regression errors are larger for small objects, and the model struggles with key feature extraction in large background interference, leading to significant performance bottlenecks in small object and blurred boundary object detection.

To enhance YOLOv8's ability to express fine-grained features, this paper introduces CBAM (Convolutional Block Attention Module), a lightweight attention mechanism module that effectively increases the model's focus on key features. CBAM consists of two modules: the channel attention and spatial attention modules, which respectively apply weighting to the feature map along the channel and spatial dimensions, allowing the model to focus more on important areas of the image, especially blurred and small objects.

CBAM first applies the channel attention module to weight the feature map. Specifically, the channel attention module computes the channel attention weight M_c by performing global average pooling and max pooling on the input feature map F_{yolov8}:

$$M_c = \sigma(W_c([\arg(F_{\text{yolov8}}), max(F_{\text{yolov8}})]))$$

whereW_c is the learnable weight, σ is the activation function, and $\text{avg}(F_{\text{yolov8}})$ and $\max(F_{\text{yolov8}})$ are the global average and max pooling operations, respectively. The channel attention mechanism helps the model identify which channels are most important for object detection, especially for blurred boundaries and small objects. The weighted feature map $F_{\text{cbam\_channel}}$ is then calculated as:

$$F_{\text{cbam}_{\text{c}}\text{hannel}} = F_{\text{yolov8}} \times M_c$$

Next, CBAM applies the spatial attention module to weight the feature map along the spatial dimension, focusing on important areas of the image. The spatial attention module computes the spatial attention weight M_s by performing average pooling and max pooling operations on the channel-weighted feature map $F_{\text{cbam\_channel}}$:

$$M_s = \sigma(W_s([\text{avg}(F_{\text{cbam}_{\text{c}}\text{hannel}}), max(F_{\text{cbam channel}})]))$$

The final spatially weighted feature map is given by:

ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

$$F_{\mathrm{cbam_final}} = F_{\mathrm{cbam_channel}} \times M_s$$

In this way, CBAM allows the network to focus more on important regions in the image, improving detection accuracy, especially for small and blurred objects.

PLPNet (Pyramid-like Localized Patch Network) further improves YOLOv8's backbone network by enhancing multi-scale feature extraction and the sensitivity to local details. PLPNet extracts different scale local patches through a pyramid structure, which effectively increases the network's sensitivity to small objects and fine-grained features.

For the input feature map $F_{\text{cbam final}}$,PLPNet extracts local patches $P_i(F_{\mathrm{cbamfinal}})$ at different scales:

$$P_i(F_{\mathrm{cbam_final}}) = \mathrm{PatchExtraction}(F_{\mathrm{cbam_final}}, \mathrm{scale}_i)$$

These local patches are extracted at different scales and weighted, resulting in the final multi-scale feature map F_{PLPNet}:

$$F_{\mathrm{PLPNet}} = \sum_{i=1}^{N} P_i(F_{\mathrm{cbam_final}})$$

where *N* is the number of pyramid layers. This method enables PLPNet to extract multi-scale features from the image, particularly enhancing fine details in small objects, which significantly improves YOLOv8's robustness.

After the improvements from CBAM and PLPNet, the final feature map F_{final} for YOLOv8 is calculated as:

$$F_{\mathrm{final}} = F_{\mathrm{PLPNet}} \times F_{\mathrm{cbam_final}}$$

This process combines the feature map weighted by CBAM with the multi-scale features from PLPNet, resulting in an enhanced feature representation that incorporates both multi-scale information and key features.

3. Dataset and Preprocessing

3.1. Dataset Description

The dataset used in this study consists of 1,659 images of BoShao (Bozhang), collected from agricultural fields. The images were manually labeled to indicate the presence of BoShao in each instance. This dataset provides the foundation for training and testing the object detection model for BoShao recognition. Due to natural variability in environmental conditions, such as lighting, background complexity, and plant growth, the images contain various challenges for accurate recognition.

3.2 Dataset Split and Statistics

To evaluate the performance of the proposed model, the dataset is divided into a training set and a test set. The training set consists of 1,493 images, accounting for 90% of the dataset, while the test set contains 166 images, representing 10% of the dataset. The total number of images in the dataset is 1,659.The dataset distribution is shown in Table 1:

Table 1: Dataset Partition and Proportions

Category	Number of Images	Proportion
Training Set	1493	90%
Test Set	166	10%
Total	1659	100%

This split ensures that a sufficient number of images are used for training, while a separate test set is

ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

available to evaluate the model's generalization capabilities on unseen data.

3.3 Data Preprocessing and Augmentation Methods

Given the specific characteristics of BoShao and its potential for appearing in different lighting conditions, growth stages, and occlusion scenarios, data augmentation plays a vital role in enhancing model performance. The augmentation techniques used in this study include:

Random Rectangular Occlusion: This technique simulates occlusion by randomly placing rectangular blocks over the image. This helps train the model to recognize BoShao even when it is partially covered, which is a common scenario in real-world applications where plants may overlap or be obstructed by environmental factors.

Gaussian Noise: Adding Gaussian noise to images helps simulate real-world imperfections like sensor noise, blurry images, or atmospheric disturbances, which improves the model's robustness to such artifacts.

Flipping: Randomly flipping the images horizontally helps expose the model to variations in object orientation, which is particularly useful when detecting BoShao from different perspectives.

Scaling: Resizing the image randomly helps the model learn to detect objects at different sizes, which is important for recognizing BoShao at various scales due to variations in plant size or camera distance.

These augmentation methods are applied as online augmentation, meaning the transformations are applied in real-time during the training process. This approach ensures that each image seen by the model is slightly different, encouraging the model to learn more generalized features and improving its performance on unseen data.

By using these four augmentation techniques, we ensure that the model is exposed to a diverse set of images, which helps it generalize better to various real-world conditions.

4. Experiments and Results

4.1 Configuration

To ensure the stability and reliability of the experiments, all tests were conducted using the same hardware and software configuration. The detailed configuration is provided in Table 2：

Table 2: Hardware and Software Configuration

Category	Configuration
Hardware	
GPU	NVIDIA A100, 80 GB memory
CPU	Intel Xeon Platinum series
Memory	256 GB
Software	
Operating System	Ubuntu 20.04 LTS
Programming Language	Python 3.9
Deep Learning Framework	PyTorch 1.10.1，Torchvision 0.11.1
CUDA Toolkit	CUDA 10.1

ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

4.2Ablation Experiments

To evaluate the effect of different model configurations, a series of ablation experiments were conducted. The experimental groups include:

Experiment 1: YOLOv8 baseline model

Experiment 2: YOLOv8 + CBAM

Experiment 3: YOLOv8 + PLPNet

Experiment 4: YOLOv8 + CBAM + PLPNet

All experiments used the same hyperparameters, as shown in Table 3:

Table 3: Hyperparameter Settings for Ablation Experiments

Category	Recommended Setting
Learning Rate	0.0005
Batch Size	16
Number of Epochs	100
Regularization	L2 Regularization 0.0005, Dropout 0.4

4.3 Ablation Experiment Analysis

This study conducted ablation experiments on YOLOv8 models with different configurations, evaluating the models' precision and recall rates, and performing a comparative analysis of the effects of various improvements. As shown in Table 4:

Table 4: Performance Comparison of Different Model Configurations in Ablation Experiments

Model	mAP@0.5	mAP@0.95	Precision	Recall
YOLOv8 (Baseline)	96.38	55.89	96.29	93.33
YOLOv8+CBAM	98.07	58.61	97.78	98.18
YOLOv8+PLPNet	98.3	59.84	97.59	98.08
YOLOv8+PLPNet+CBAM	98.66	60.7	98.69	98.18

Single Module Improvement: Adding either CBAM or PLPNet alone improves the baseline model's performance. PLPNet shows a more significant improvement in mAP@0.95 (+3.95% vs. CBAM's +2.72%), indicating that multi-scale feature fusion is more critical for high IoU threshold detection.

Synergistic Effect: When CBAM and PLPNet are used together, the model achieves the best results in mAP@0.5 (98.66%), precision (98.69%), and recall (98.18%), confirming the complementarity of the attention mechanism and multi-scale learning.

Recall rate refers to the proportion of correctly identified positive samples among all the true positive samples detected by the model. The changes in recall rate at various thresholds for the YOLOv8 baseline model, YOLOv8 + CBAM, YOLOv8 + PLPNet, and YOLOv8 + CBAM + PLPNet are shown in Figure 1:

ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

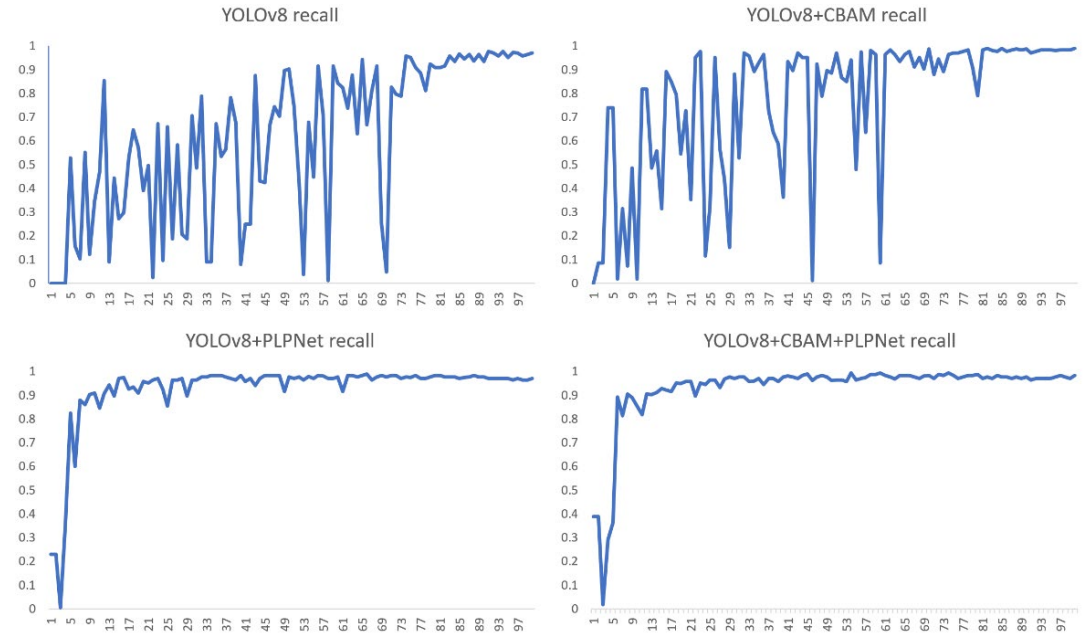

Figure 1: Recall Rate Curves Across Different Model Configurations

The baseline model (YOLOv8) shows significant recall fluctuation in the low-threshold range, indicating poor consistency in detecting blurred targets.

After adding CBAM (YOLOv8 + CBAM), the recall in the high-threshold range significantly improves (+4.85%), validating the attention mechanism's enhancement of key features.

PLPNet (YOLOv8 + PLPNet) performs steadily in the mid-high threshold range, demonstrating the adaptability of multi-scale learning to complex backgrounds.

The combined model (YOLOv8 + CBAM +PLPNet) maintains a high recall across the entire threshold range, proving the modules' synergy in effectively reducing missed detections.

Precision reflects the proportion of correctly identified targets among all the positive samples detected by the model. The changes in precision at various thresholds for the YOLOv8 baseline model, YOLOv8 + CBAM, YOLOv8 + PLPNet, and YOLOv8 + CBAM + PLPNet are shown in Figure 2:

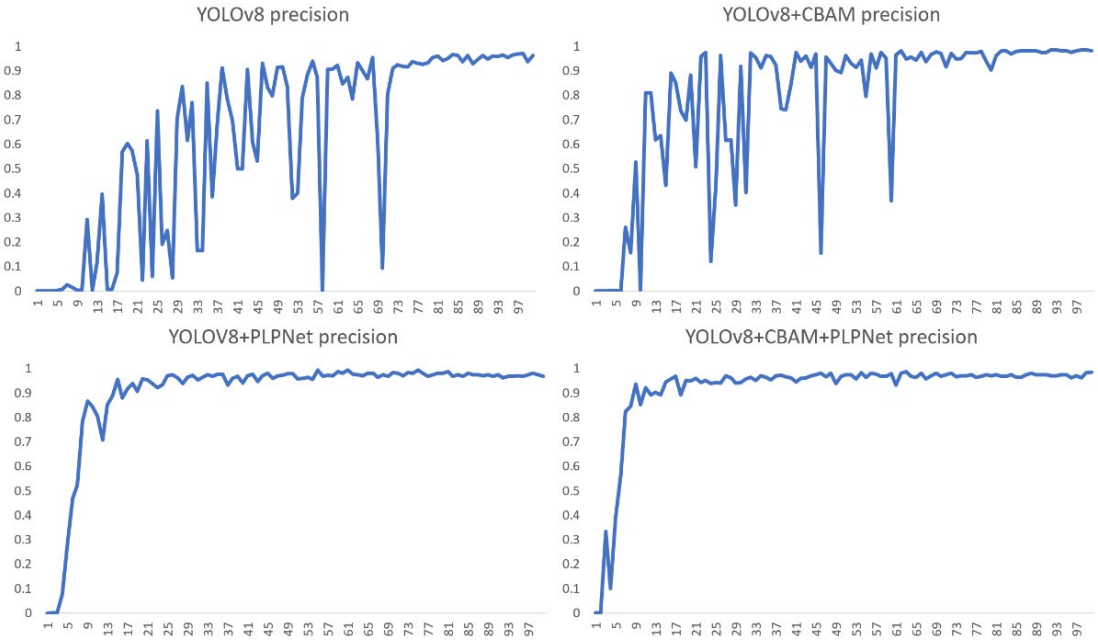

Figure 2: Precision Curves Across Different Model Configurations

The baseline model (YOLOv8) shows fluctuations in precision in the low-threshold range, reflecting its inadequate resistance to background interference.

ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

CBAM (YOLOv8 + CBAM) improves high-threshold precision to 97.78% by suppressing irrelevant features.

PLPNet (YOLOv8 + PLPNet) optimizes localization accuracy through multi-scale fusion, reducing false detections caused by size variations.

The combined model (YOLOv8 + CBAM + PLPNet) exhibits a smooth and overall high precision curve, indicating stable detection performance across different confidence thresholds.

CBAM and PLPNet address key bottlenecks in small object detection through feature selection enhancement and multi-scale contextual awareness, respectively. Their synergistic effect optimizes the model's overall performance (mAP@0.5 increased by 2.28%) and robustness (mAP@0.95 increased by 4.81%) in complex scenarios.

4.4 Comparison Experiments

To comprehensively evaluate the performance of the improved YOLOv8 model in the BoShao detection task, this study conducted comparison experiments with several advanced object detection models. The models selected for comparison include **Gold-YOLO**, **YOLOv11**, **EfficientDet**, **RetinaNet**, and **CenterNet**.

The experimental parameters and environmental configurations were the same as those used in the previous experiments.

In the experiments, the models were evaluated based on **mAP@0.5**, **mAP@0.95** (mean average precision at a higher IoU threshold), **Precision**, and **Recall** to assess their performance in the BoShao detection task. The results are presented in Table 5:

Table 5: Performance Comparison with State-of-the-Art Models

Model	mAP@0.5	mAP@0.95	Precision (p)	Recall (R)
Gold-YOLO	96.14	56.88	96.42	94.12
YOLOv11	97.83	59.32	97.9	97.15
EfficientDet	97.56	58.21	97.61	96.83
RetinaNet	96.87	57.43	96.8	95.55
CenterNet	94.89	54.13	94.67	93.12
YOLOv8+CBAM+PLPNet	98.66	60.7	98.69	98.18

As shown in Table 5, the combination of **YOLOv8 + CBAM + PLPNet** demonstrates exceptional performance in the BoShao detection task, achieving the best results in **mAP@0.5**, **mAP@0.95**, and **Recall**.Specifically, **YOLOv8 + CBAM + PLPNet** achieved **98.66** in **mAP@0.5** and **60.70** in **mAP@0.95**, showing remarkable precision at higher IoU thresholds, indicating that this model can accurately identify targets while reducing errors in detection.Additionally, the **Precision** of **YOLOv8 + CBAM + PLPNet** is **98.69**, outperforming all other models and demonstrating its ability to detect targets with high accuracy while effectively reducing false positives.For **Recall**, the model achieved **98.18**, indicating that **YOLOv8 + CBAM + PLPNet** excels at capturing and detecting more BoShao targets, particularly in complex backgrounds and small object detection, with a significant reduction in false negatives.

The comparison experiments led to the following conclusions:

Best Performance of YOLOv8 + CBAM + PLPNet: This combination shows the best performance in mAP@0.5, mAP@0.95, Precision, and Recall for BoShao detection, confirming that the improvements of CBAM and PLPNet significantly enhance YOLOv8's performance in BoShao detection.

Performance Advantage: YOLOv8 + CBAM + PLPNet outperforms other models in the BoShao detection task, particularly in mAP@0.5 and mAP@0.95, demonstrating its superior performance in complex scenes and small object detection.

In summary, the combination of **YOLOv8 + CBAM + PLPNet** performs excellently in the BoShao detection task, significantly improving detection precision, recall, and robustness, making it highly

ISSN 2616-5775 Vol. 8, Issue 3: 28-35, DOI: 10.25236/AJCIS.2025.080304

valuable for practical applications

5. Conclusion

In this study, a novel approach to BoShao (Bozhang) recognition was proposed by integrating YOLOv8 with the Convolutional Block Attention Module (CBAM) and Pyramid-like Localized Patch Network (PLPNet). The results demonstrate significant improvements in detection accuracy and robustness, particularly in handling small objects, blurred boundaries, and complex backgrounds, which are common challenges in BoShao recognition tasks.

The ablation experiments revealed that the addition of CBAM and PLPNet to YOLOv8 enhanced both the recall and precision, with the best results achieved by the YOLOv8 + CBAM + PLPNet model. This combination outperformed the baseline YOLOv8 model and demonstrated superior performance in terms of mean average precision (mAP) at both IoU thresholds (mAP@0.5 and mAP@0.95), as well as recall and precision. Notably, the integration of CBAM allowed the model to focus on key image features, and PLPNet improved multi-scale feature extraction, addressing the limitations of YOLOv8 in detecting small objects and dealing with complex environmental conditions.

Moreover, the comparison experiments with other advanced object detection models such as YOLOv11, EfficientDet, RetinaNet, and CenterNet further confirmed the efficacy of the proposed model. YOLOv8 + CBAM + PLPNet outperformed these models in key metrics, particularly in scenarios involving small object detection and cluttered backgrounds, which are frequently encountered in real-world BoShao recognition applications.

In conclusion, the combination of YOLOv8, CBAM, and PLPNet provides a robust solution for BoShao recognition, offering substantial improvements over previous methods. This approach demonstrates not only enhanced detection accuracy and reliability but also the potential for real-world applications in the automated identification of BoShao in agricultural settings. The findings underline the importance of integrating attention mechanisms and multi-scale feature extraction techniques to address the challenges posed by complex and varied environments in object detection tasks.

References

[1] Ultralytics. (2024). YOLOv8: A new version of YOLO for object detection. GitHub. https://github.com/ultralytics/yolov8
[2] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), 3–19.
[3] Tang, Z., He, X., Zhou, G., Chen, A., Wang, Y., Li, L., & Hu, Y. (2023). A Precise Image-Based Tomato Leaf Disease Detection Approach Using PLPNet. Plant Phenomics, 5, 0042.