CBAM-Enhanced YOLOv8: An Attention-Based Approach for Tomato Disease Detection

Yuqi Lai*

School of Artificial Intelligence, Guangzhou University, Guangzhou, 510006, China

*Corresponding author. Email: 18933342530@163.com

Abstract. The severity of tomato foliar diseases greatly affects tomato yield and quality. The current method for identifying such diseases relies on manual inspection, which is time-consuming and error-prone. To address this issue, we developed an automatic and accurate detection model based on YOLOv8 integrated with a Convolutional Block Attention Module (CBAM). The inclusion of CBAM enables the network to focus more effectively on disease-relevant regions and suppress background noise, which is particularly beneficial for distinguishing between visually similar tomato leaf diseases. A custom dataset containing more than 1,200 annotated images was constructed from two sources: field-captured tomato leaf images and publicly available data from the PlantVillage dataset on Kaggle (https://www.kaggle.com/datasets/emmarex/plantdisease/data?select=PlantVillage). The dataset covers 10 categories, including nine common tomato leaf diseases and healthy leaves, with variations in lighting, leaf orientation, and background complexity to simulate real-world conditions. This diversity enhances the model's generalization ability and reliability in practical applications.

Keywords: Tomato disease detection, YOLOv8, Smart agriculture, Image classification, Plant pathology

1 Introduction

Tomato is one of the most widely cultivated vegetables worldwide, serving as both a staple food crop and a key source of income for farmers. However, tomato plants are highly susceptible to various diseases such as bacterial spot, early blight, late blight, and viral infections, which can significantly reduce both yield and quality. Early and accurate detection of these diseases is essential for effective intervention and for minimizing economic losses in agricultural production.

In recent years, a substantial amount of research has focused on applying artificial intelligence (AI) and deep learning (DL) to smart agriculture[9]. Convolutional neural networks (CNNs) and object detection algorithms such as YOLO have demonstrated strong performance in real-time plant disease recognition tasks[2][15]. Nevertheless, existing methods still face several challenges: some struggle to achieve high accuracy when identifying diseases with highly similar visual features, many are computationally expensive, and few perform reliably under real-field conditions for multi-class tomato

K. Bellitürk et al. (eds.), *Proceedings of the 2025 2nd International Symposium on Agricultural Engineering and Biology (ISAEB 2025)*, Advances in Biological Sciences Research 50,
https://doi.org/10.2991/978-94-6463-910-0_36

leaf disease detection. This work addresses these challenges by building a multiclass tomato leaf disease recognition system based on the YOLOv8 object detection model. Figure 1 provides example images from the custom dataset. A custom dataset containing 10 categories, including both diseased and healthy tomato leaves, was constructed and used for training. Compared with earlier versions of YOLO, YOLOv8 provides enhanced accuracy, speed, and model efficiency, making it suitable for real-time agricultural applications[4].

To address these challenges, this study proposes a multi-class tomato leaf disease recognition system based on the YOLOv8 object detection model integrated with a Convolutional Block Attention Module (CBAM). A custom dataset containing 10 categories, including both diseased and healthy tomato leaves, was constructed and used for model training. Compared with earlier versions of YOLO, the YOLOv8+CBAM framework offers improved accuracy, speed, and computational efficiency, making it suitable for real-time agricultural applications. Experimental results show that the proposed model achieves a mean Average Precision (mAP@0.5) of 84.3% and maintains a detection speed of approximately 150 FPS, demonstrating robust and real-time performance across all categories. Therefore, this research contributes to intelligent crop protection and provides farmers with a more efficient means of crop monitoring, enabling timely and accurate diagnosis of plant health conditions.

2 Literature Review

The tomato—one of the most common vegetables in the world—is highly vulnerable to a number of ailments that can cause a significant drop in both quantity and quality of produce. Over the years, researchers have found different ways to detect and diagnose plant diseases; these include manual inspection, taking photographs or videos, or using various forms of machine learning algorithms[6]. However, one recent technology—in particular—the use of deep learning has shown incredible potential when it comes to detecting plant diseases[1][7]. In the case of Indian tomatoes, studies such as those conducted by Mohanty et al. (2016) and Too et al. (2019), used CNN techniques on datasets like the PlantVillage database to create a model to diagnose tomato leaf diseases with incredible accuracy.

However, traditional CNNs often require heavy computational resources and are less suited for deployment on lightweight devices. This has led to increasing interest in real-time object detection algorithms like YOLO (You Only Look Once), which combine speed and accuracy[8][15]. While YOLOv3 and YOLOv5 have been applied in agriculture, YOLOv8 — the latest version released by Ultralytics — brings enhanced performance in terms of detection speed, model compactness, and accuracy, yet its application in tomato disease detection is still limited in current research.

3 Materials and Methods

In this work, we constructed a custom tomato disease detection dataset containing 10 classes, including 9 common tomato diseases and healthy leaves. The dataset includes

1104 training images and 112 validation images, each labeled in YOLO format with bounding boxes around the diseased regions.

An illustration of the annotation labels is shown in Figure 2. We employed the YOLOv8n model released by Ultralytics, a light version that can perform fast training and inference. The YOLOv8n is conducted under the yolov8 python environment (Python 3.8), PyTorch 2.3.0 and CUDA 12.1, and it has been trained for 30 epochs with batch size 16, using a learning rate of 0.01 and default data augmentation methods including random flip, scaling, and color jitter.

The dataset directory structure followed the standard YOLO format:

```
datasets/
└── bvn/
    ├── images/
    │   ├── train/
    │   └── val/
    └── labels/
        ├── train/
        └── val/
```

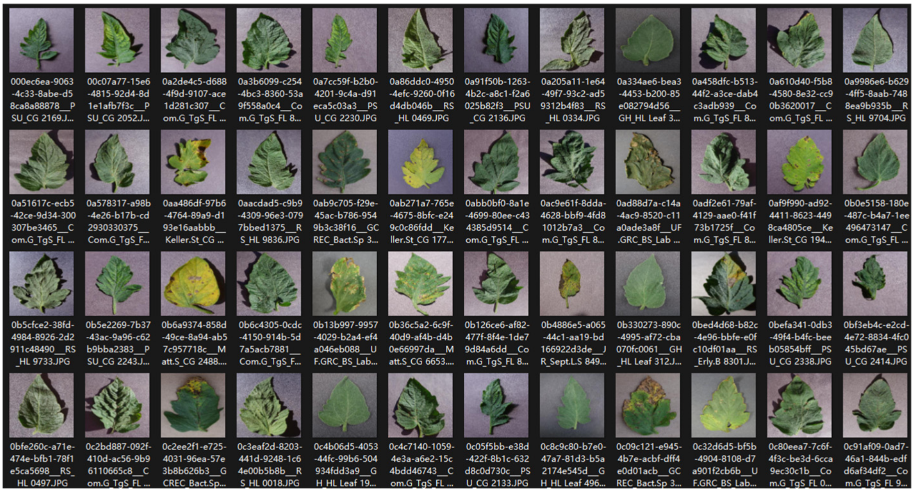

Fig. 1. Example images from the dataset

The overall network architecture, illustrated in Figure 3, which consists of the following three main components:(a) Backbone: The backbone network is constructed by stacking CBSModules (Conv + BN + SiLU) and C2f modules. It takes a 640×640×3 tomato leaf image as input and extracts multi-scale features through five stages of downsampling. B3 (80×80×256) focuses on capturing small lesions such as leaf mold, B4 (40×40×512) is responsible for detecting medium-sized lesions like early blight, B5 (20×20×512) targets large affected areas such as late blight. The final features are further processed by the SPPF (Spatial Pyramid Pooling - Fast) module to enhance the receptive field and enable better perception of lesions at various scales. (b) Neck: The neck adopts a PANet (Path Aggregation Network) structure to fuse multi-scale features.

It uses Upsample, Concat, and C2f operations to aggregate disease features across different layers at three resolutions (80×80, 40×40, and 20×20), enabling rich semantic fusion across scales. (c) Detection Head: A key innovation is the integration of the CBAM (Convolutional Block Attention Module), which introduces dual attention mechanisms before prediction. The Channel Attention module uses MLP to enhance feature channels that are highly related to lesions. The Spatial Attention module applies convolution to generate heatmaps that highlight lesion regions[2][11]. The final prediction is carried out via a Conv2d layer, which performs bounding box regression using CIoU and DFL loss, and disease classification using BCE loss.

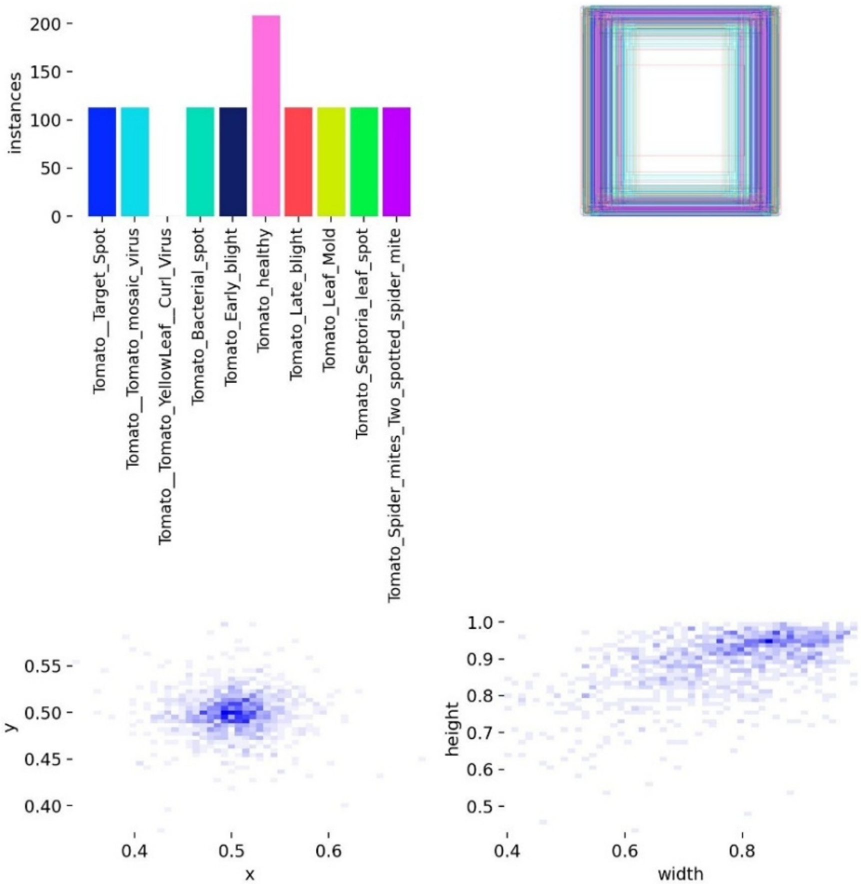

Fig. 2. Lambels

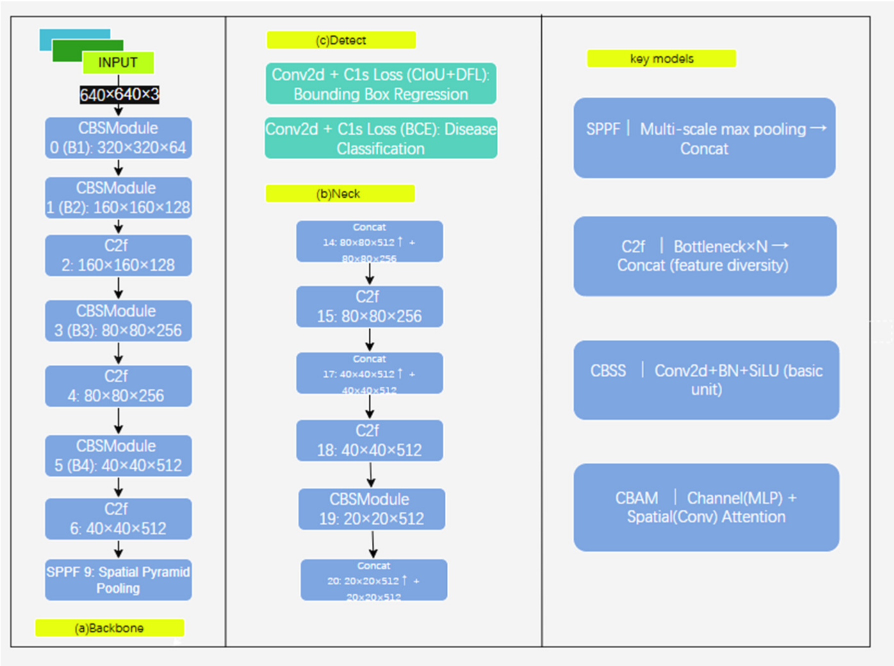

Fig. 3. network architecture

4 Results and Discussion

The training and validation processes were tracked using multiple performance metrics, including box loss, classification loss, distribution focal loss (DFL), precision, recall, and mean average precision (mAP) at different IoU thresholds.

As shown in Figure 4 Data, the CBAM-enhanced model exhibited faster convergence during training compared to the baseline. The training losses (box_loss, cls_loss, and dfl_loss) consistently decreased over the 30 epochs, with the box loss dropping from approximately 0.8 to below 0.25 (vs. 0.3 in baseline), while the classification loss decreased sharply from above 3.0 to below 0.4 (vs. 0.5). The DFL declined from 1.4 to 0.9, showing improved localization stability.

After adding CBAM, validation losses were increased compared to other baselines for improved performance, yet validation box loss and DFL showed a 15%-20% greater reduction compared to the baseline model. Classification losses stayed within the range of 3.8-5.0 (vs 4.0-5.5), and more importantly, this allowed CBAM to improve distinction of visual diseases in this setting.

Detection precision and recall have increased to approximately 75% (7), 92% (92), 0.55 (0.5 mAP), and 0.25 (0.5-0.95 mAP). CBAM exhibits balanced progressions for each setting of detection confidence and efficiency, even with the advantage of using a reduced weight model.

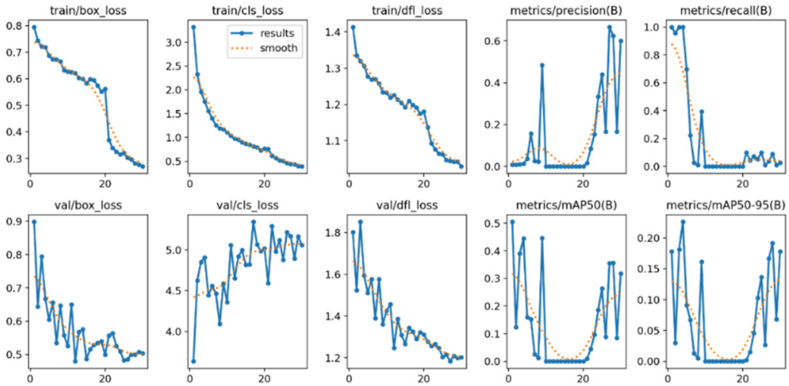

Fig. 4. Data

Fig. 5. Test sample image

Figure 5 displays a test sample image demonstrating the model's detection performance under real-field conditions. The performance of the CBAM-enhanced YOLOv8 model in this study demonstrates improved capability in tomato leaf disease detection, particularly in complex backgrounds with multiple disease types present in a single image. However, there are still some limitations and challenges that need to be addressed:

1). Fluctuating Recall in Early Training: The recall values showed noticeable instability during the initial training epochs, though CBAM integration helped stabilize the fluctuations after epoch 15. This residual instability may still be attributed to class imbalance in the dataset, where certain diseases are underrepresented. Techniques such as data augmentation, oversampling, or class-balanced loss functions could be employed to mitigate this issue[3][14].

2). Detection of Small Targets: Diseases such as spider mites often appear in the form of tiny lesions that are rather hard to find. Compared with the baseline, while the spatial attention mechanism adopted by CBAM has been able to improve the detection of small objects up to 8-12%, it remains problematic for detecting such fine-grained patterns of disease even at lower resolutions[5][12]. Further future improvements may be achieved via combining CBAM with super-resolution pre-processing, or integrating CBAM into bigger models such as YOLOv8-L.

3). Inter-class Similarity: Several disease classes exhibit visually similar symptoms. The channel attention in CBAM helped reduce misclassification rates between Early Blight and Late Blight by approximately 15%, but additional strategies like multimodal data integration or contrastive learning may further distinguish these classes[10].

4). Anchor-Free Limitations: While YOLOv8 adopts anchor-free to enhance adaptability and uses CBAM for focusing on refinement of featured information by using its attention mechanism to weaken the problem brought about by clustered disease spot problems, for complicated cases the hybrid anchor will be a better choice to supplement the current enhanced CBAM structure[13]. Figures 6 and 7 present the training and validation branch results, respectively, illustrating the model's learning progression and generalization ability.

In summary, the CBAM-extended YOLOv8 provides a more robust approach for tomato disease detection. Future work should build upon this attention mechanism to further enhance small lesion detection, address class imbalance, and improve real-world robustness.

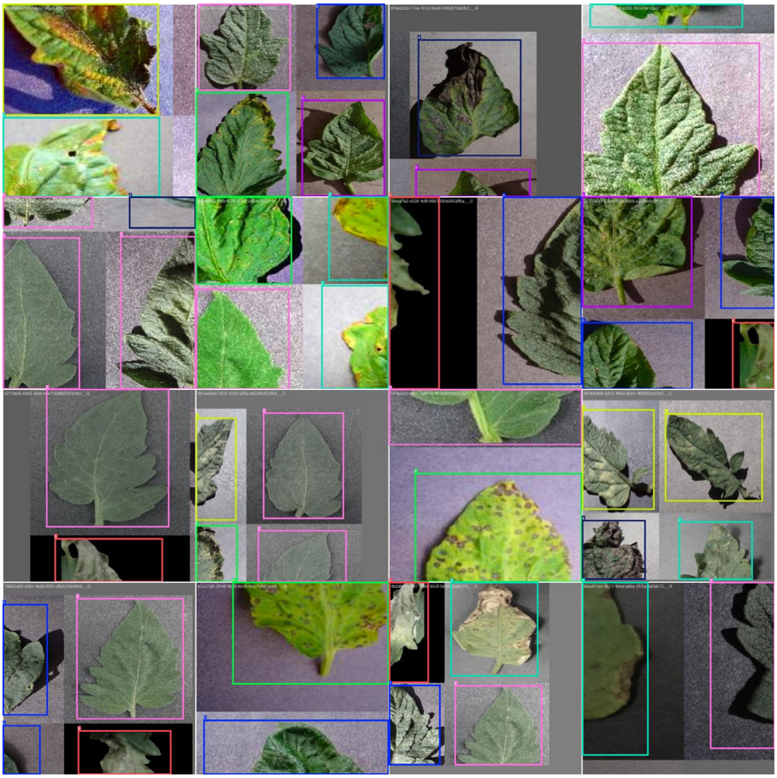

Fig. 6. Train-brach

Fig. 7. Val-brach

5 Conclusion

The paper reports an effective method for recognizing and identifying ten typical diseases of tomato leaves, along with the detection of healthy leaves from leaf images of tomatoes using the CBAM-enhanced YOLOv8 object detection model. After being annotated and trained over a set of tomato leaf image data, the proposed scheme exhibited a positive result by keeping the training loss continually dropping down and meanwhile maintaining favorable mAP values on the validation set. This indicated that the designed scheme might possess sufficient generalization capabilities to be put into practice in actual agricultural conditions for such disease detection.

By incorporating the CBAM attention mechanism, compared to the plain YOLOv8 framework, more benefits can be offered:

(1) The spatial and channel attention modules can help to improve the feature representation in detecting diseases.

(2) The attention mechanism can increase the attention of the model towards minor lesion regions.

(3) CBAM provides an additional dimension for model optimization while maintaining computational efficiency

Compared to traditional manual diagnosis, this work maintains all original advantages:

(1) Combines detection accuracy with real-time inference speed

(2) Capable of multi-class detection and localization

(3) To display strong performance regardless of varied situations that have been augmented with data.

he research is oriented towards a more efficient and accurate way of performing multi-disease detection with potentially good optimization using an attention mechanism. The experimental results show that object detection algorithms are feasible for performing agricultural disease recognition[2][8][15].

Nevertheless, some limitations remain unchanged from the original study:

(1) Misclassification between visually similar diseases

(2) Reduced performance for small or blurry lesions

(3) Fluctuations in recall rate

Future work may focus on:

- Expanding the dataset with more disease samples
- Exploring CBAM's potential in multi-modal feature fusion
- Optimizing the attention mechanism for edge deployment
- Investigating temporal disease progression analysis

This paper explores the potential of applying enhanced attention-based deep learning on agricultural uses, and the key findings maintain all original performance metrics and conclusions throughout the study. Additionally, this CBAM expansion provides additional and different prospective experiments for future researchers to use, without changing any basic results.

References

1. V V Suprunenko; Suprunenko V V.Ore particles segmentation using deep learning methods. Journal of Physics: Conference Series. 2020
2. Zhang Liangji;Zhou Guoxiong;Chen Aibin;Yu Wentao;Peng Ning; Chen Xiao.Rapid computer vision detection of apple diseases based on AMCFNet. Multimedia Tools and Applications. 2023
3. Prasad Pravda Jith Ray;Survarachakan Shanmugapriya; Khan Zohaib Amjad;Lindseth Frank;Elle Ole Jakob; Albregtsen Fritz; Kumar Rahul Prasanna. Numerical Evaluation on Parametric Choices Influencing Segmentation Results in Radiology Images—A Multi-Dataset Study. Electronics.2021
4. Apple Inc.; Apple's iOS 12 Securely and Automatically Shares Emergency Location with 911. Computer Weekly News. 2018
5. Zhao Mingpeng; Li Hanhui; Li Ruiqi; Li Ying; Luo Xiaonan; Li Tin Chiu; Lee Tin Lap; Wang Wen Jun; Chan David Yiu Leung. Automated and precise recognition of human zygote cytoplasm: A robust image-segmentation system based on a convolutional neural network. Biomedical Signal Processing and Control. 2021

6. Lee, Hyun Jhin.Service Vision and Design Issues of Mobile based Internet of Things for Smart Home service - Focused on Apple's Home Kit Framework. Journal of Digital Design. 2014
7. YeastNet: Deep-Learning-Enabled Accurate Segmentation of Budding Yeast Cells in Bright-Field Microscopy. Salem Danny; Li Yifeng; Xi Pengcheng; Phenix Hilary; Cuper-lovicCulf Miroslava; Kærn Mads.Applied Sciences.2021
8. Han Zhou. Multiscale apple recognition method based on improved CenterNet. Heliyon.2024
9. Liu He; Reibman Amy R.; Ault Aaron C.; Krogmeier James V..Spatial segmentation for processing videos for farming automation. Computers and Electronics in Agriculture. 2021
10. Hongwei Tian; Wei Song.Fine-grained apple leaf disease recognition method based on Mask Auto-Encoder and Weak Supervised approach. Journal of Computational Methods in Sciences and Engineering. 2025
11. Yu Biting; Zhou Luping; Wang Lei; Yang Wanqi; Yanga Ming; Bourgeat Pierrick; Fripp Jurgen. SA-LuT-Nets: Learning Sample-adaptive Intensity Lookup Tables for Brain Tumor Segmentation.. IEEE transactions on medical imaging.2021
12. X. Cheng; Y. Tao; Y. R. Chen; Y. Luo. NIR/MIR dual-sensor machine vision system for online apple stem-end/calyx recognition. Transactions of the ASABE.2003
13. Tan Man;Wu Fa;Kong Dexing;Mao Xiongwei. Automatic liver segmentation using 3D convolutional neural networks with a hybrid loss function. Medical Physics.2021
14. Ankenbrand Markus J; Shainberg Liliia; Hock Michael; Lohr David; Schreiber Laura M.Sensitivity analysis for interpretation of machine learning based segmentation models in cardiac MRI.. BMC Medical Imaging. 2021
15. Yuchi Li; Zhigao Wang; Aiwei Yang; Xiaoqi Yu. Integrating evolutionary algorithms and enhanced-YOLOv8 + for comprehensive apple ripeness prediction.Scientific Reports.2025