

“Attention-Based Multi-Scale Deep Learning for Tomato Leaf Disease Detection in Complex Natural Environments”

Dr. Shide J. U.

^{1*}Department of Botany, Shivneri Mahavidyalaya Shirur Anantpal Dist. Latur (MS) India 413544

Email:- dr.jayashri713@gmail.com

Abstract

Tomato is an economically important vegetable crop cultivated extensively in different parts of the world. However, its production is seriously affected by fungal, bacterial, viral, and insect-related diseases. Early identification of disease symptoms is essential for reducing disease spread, minimizing crop losses, and supporting timely crop-management decisions. Conventional disease identification largely depends on visual inspection and expert knowledge, which can be time-consuming and difficult to apply over large cultivated areas. Recent advances in deep learning and computer vision have provided effective approaches for automatic plant disease detection. Nevertheless, disease detection in natural agricultural environments remains challenging because of variations in illumination, shadows, leaf occlusion, complex backgrounds, overlapping leaves, and small or visually similar lesions. The present research paper examines an attention-based and multi-scale feature-fusion approach for tomato leaf disease detection using an improved YOLOv6 architecture. The method integrates the Convolutional Block Attention Module (CBAM) into the backbone network and introduces a BiRepGFPN feature-fusion module incorporating shallow P2 features. The approach was evaluated on a tomato leaf disease dataset containing 1,425 images representing tomato septoria leaf spot, early blight, leaf mold, late blight, leaf miner, and healthy leaves. Data augmentation increased the training images to 4,659. The improved model achieved 92.9% precision, 95.2% recall, 94.0% F1-score, and 93.8% mAP on the tomato disease dataset. It also demonstrated improved generalization on PlantDoc. The findings indicate that combining attention mechanisms with multi-scale feature fusion can substantially improve the detection of disease symptoms, particularly small lesions, in complex field conditions.

Keywords: Tomato disease, deep learning, YOLOv6, CBAM, BiRepGFPN, attention mechanism, multi-scale feature fusion, plant disease detection

1. Introduction

Tomato (*Solanum lycopersicum* L.) is one of the most widely cultivated vegetable crops and represents an important source of food, nutrition, and income. Tomato plants are vulnerable to a variety of diseases that can reduce plant growth, fruit quality, and overall yield. Disease symptoms frequently appear first on leaves, making leaf examination an important component of disease diagnosis. Early and accurate identification is therefore essential for effective disease management. The uploaded study emphasizes that environmental and climatic factors increase the susceptibility of tomatoes to fungal, bacterial, and viral infections and that early disease recognition can help reduce infection and spread.

Traditional plant disease diagnosis is generally based on visual examination by farmers, agricultural officers, or plant pathology experts. Although experienced experts can recognize many symptoms, conventional inspection becomes difficult when large numbers of plants have to be monitored. It can also be affected by observer experience, fatigue, environmental conditions, and similarity among disease symptoms. Deep learning-based computer vision has consequently emerged as an important technology for automated disease detection. Such systems can learn disease-related visual features directly from images and can potentially provide rapid and consistent disease recognition.

Several object-detection frameworks, including Faster R-CNN, SSD, and the YOLO family, have been applied to agricultural image analysis. The YOLO family is particularly attractive because it provides a useful balance between detection accuracy and computational speed. The source study selected YOLOv6 as its baseline because of its suitability for real-time object detection. However, conventional YOLOv6 has limitations when disease symptoms are very small or when images contain complex environmental interference such as shadows, illumination changes, and occlusion.

Disease detection in natural environments differs considerably from detection using laboratory images. Images collected under controlled conditions generally have relatively uniform illumination and simple backgrounds. Field images, in contrast, may contain soil, weeds, shadows, overlapping leaves, variable sunlight, and other visual objects. Disease symptoms may also vary according to disease-development stage. Early blight and septoria leaf spot, for example, can produce small and visually similar circular lesions. Tomato leaf miner symptoms may resemble natural leaf textures, while leaf mold and late blight can be strongly influenced by illumination.

Attention mechanisms provide one possible solution to this problem. They allow neural networks to assign greater importance to informative features while reducing the influence of irrelevant information. The Convolutional Block Attention Module (CBAM) combines channel attention and spatial attention and can therefore emphasize both the type of feature and its spatial location. In the present approach, CBAM is integrated into the YOLOv6 backbone to enhance disease-lesion features and suppress irrelevant background information.

A second challenge is the detection of small lesions. Conventional feature-pyramid structures may lose detailed spatial information as images pass through deeper network layers. Shallow feature maps retain relatively detailed spatial information, whereas deeper feature maps contain stronger semantic information. Combining these two forms of information may improve the localization of small disease symptoms. Therefore, the source study introduces a new BiRepGFPN module by incorporating the P2 shallow feature map into the RepGFPN architecture.

The objective of this paper is to examine the effectiveness of combining CBAM attention and multi-scale BiRepGFPN feature fusion with YOLOv6 for tomato leaf disease detection under complex natural conditions.

2. Methodology

2.1 Dataset and Disease Categories

The study used a tomato leaf disease dataset assembled from publicly available sources and additional images. The dataset incorporated material from PlantVillage, PlantDoc, internet-captured images, and photographs collected by the researchers. PlantVillage contains images of several crops, including tomato, whereas PlantDoc provides images captured under natural environmental conditions.

For the investigation, six categories were considered: tomato septoria leaf spot (TSLs), tomato early blight (TEB), tomato leaf mold (TLM), tomato late blight (TLB), tomato leaf miner (TLMR), and healthy tomato leaves. The dataset consisted of 1,425 images. The original images were divided into training and testing sets in an 80:20 ratio, with the two sets maintained separately to avoid overlap.

The dataset was specifically designed to reflect difficult field conditions. It included variations in illumination, shadows, external occlusion, overlapping leaves, and differences in lesion size, shape, texture, and color. Such characteristics make the dataset more representative of practical agricultural conditions than datasets obtained exclusively under laboratory conditions.

2.2 Image Preprocessing and Augmentation

Data augmentation was applied to increase the diversity of training images. The techniques included random noise addition, brightness modification, and image flipping. After augmentation, the training dataset increased to 4,659 images.

The use of augmentation is particularly important in plant disease recognition because symptoms can appear under different illumination levels, orientations, and backgrounds. Increasing image diversity can reduce overfitting and improve the robustness of the trained model.

2.3 YOLOv6 Baseline Model

YOLOv6 is an object-detection architecture consisting of an input layer, backbone network, feature-fusion network, and detection head. Its backbone uses efficient re-parameterizable convolutional structures to extract features, while its feature-fusion network combines information at multiple scales. The original YOLOv6 uses a Path Aggregation Feature

Pyramid Network (PAFPN). Feature maps corresponding to different spatial scales are used for detecting small, medium, and large objects.

Although this structure is effective for general object detection, disease lesions in natural tomato images can be substantially smaller and less distinctive than ordinary objects. Consequently, improvements are required to preserve detailed spatial information and suppress irrelevant environmental information.

2.4 Integration of CBAM Attention Mechanism

The Convolutional Block Attention Module was incorporated into the backbone after three CSPRepBlocks. CBAM consists of channel and spatial attention components.

The channel attention mechanism determines which feature channels are important. Average pooling and maximum pooling are applied to the feature map, and the resulting information is processed to produce a channel attention map. The spatial attention mechanism subsequently identifies important spatial locations by applying average and maximum pooling followed by convolution.

The combined attention operation enables the model to concentrate on disease-related regions while reducing the influence of background information. The source study reports that CBAM improved the model's ability to focus on tomato disease features under complex backgrounds.

2.5 BiRepGFPN Multi-Scale Feature Fusion

The second major modification involved replacing the original PAFPN with BiRepGFPN. Feature Pyramid Networks are designed to combine information from different feature levels. However, ordinary feature fusion may not sufficiently preserve fine spatial information required for identifying small lesions.

RepGFPN provides more efficient feature aggregation through cross-scale connections and re-parameterized structures. The proposed BiRepGFPN further introduces the P2 feature map, which contains relatively detailed spatial information. The P2 feature is combined with other feature levels to improve localization of small lesions.

The combination of shallow and deep features is conceptually important. Shallow features preserve boundaries, edges, and fine textures, whereas deep features provide higher-level semantic information. Their integration therefore provides a more complete representation of disease symptoms.

2.6 Training and Evaluation

The experiments were conducted using an NVIDIA RTX A6000 GPU with 48 GB VRAM, CUDA 11.1, Ubuntu 18.04.5 LTS, and PyTorch 1.9.0. The initial learning rate was 0.01, the batch size was 16, and training was performed for up to 300 epochs using the SGD optimizer. Mixed-precision training and early stopping were employed to reduce computational requirements.

Model performance was evaluated using precision, recall, F1-score, average precision (AP50), and mean average precision (mAP). Precision measures the proportion of correctly predicted positive samples among all positive predictions, whereas recall represents the proportion of actual positive samples correctly detected. The F1-score provides a combined measure of precision and recall.

3. Results and Discussion

3.1 Effect of Attention Mechanisms

Different attention mechanisms were compared, including CBAM, Efficient Channel Attention (ECA), Coordinate Attention (CA), and Efficient Multi-Scale Attention (EMA). The results demonstrated that attention mechanisms generally improved YOLOv6 performance. Among the tested mechanisms, CBAM produced the strongest overall performance in precision, recall, F1-score, and mAP.

The superiority of CBAM can be explained by its combined consideration of channel and spatial information. ECA and CA primarily emphasize channel-level information, whereas CBAM additionally identifies important spatial regions.

This characteristic is particularly useful for disease detection because a disease symptom is defined not only by its visual characteristics but also by its location on the leaf.

Visualization of feature maps further demonstrated that CBAM reduced attention to irrelevant background regions and increased attention to disease-related areas. The improvement was evident for early blight, late blight, leaf miner, and leaf mold. However, septoria leaf spot remained difficult because its symptoms can be small, dense, and visually indistinct.

3.2 Effect of BiRepGFPN

Feature-fusion experiments compared PAFPN, BiPAFPN, RepGFPN, and BiRepGFPN. The introduction of multi-path feature fusion improved the interaction between feature maps at different scales. The addition of the P2 feature map was particularly useful for small-lesion detection.

BiRepGFPN achieved an average precision of 99.6% for tomato leaf miner. It also improved average precision for tomato leaf mold and early blight compared with PAFPN. Although BiRepGFPN showed a slight reduction compared with RepGFPN for late blight, it still performed better than PAFPN overall.

These findings indicate that multi-scale feature fusion is especially valuable when symptoms differ considerably in size. Small lesions can disappear or become less distinguishable during repeated downsampling. Retaining P2-level information helps preserve fine spatial details and improves localization.

3.3 Ablation Analysis

Ablation experiments were performed to determine the individual and combined contributions of CBAM and BiRepGFPN. The baseline YOLOv6 model achieved 90.6% precision, 91.2% recall, 90.9% F1-score, and 91.1% mAP.

After integrating CBAM, precision increased to 90.8%, recall to 95.0%, F1-score to 92.9%, and mAP to 92.5%. When BiRepGFPN was introduced, precision was 90.1%, recall reached 97.0%, F1-score was 93.4%, and mAP was 92.6%. The complete model combining both modifications achieved 92.9% precision, 95.2% recall, 94.0% F1-score, and 93.8% mAP.

The results demonstrate that the two modifications contribute in complementary ways. CBAM is effective in suppressing environmental interference and improving feature selection, while BiRepGFPN is particularly beneficial for retaining fine spatial information and increasing recall. Combining them therefore produces a more balanced detection system.

3.4 Comparison with Other YOLO Models

The proposed model was compared with YOLOX, YOLOv5, YOLOv6, YOLOv6-s, YOLOv7, and YOLOv8. On the tomato disease dataset, the proposed method achieved 92.9% precision, 95.2% recall, 94.0% F1-score, and 93.8% mAP. The corresponding mAP values were 90.5% for YOLOX, 85.6% for YOLOv5, 91.1% for YOLOv6, 89.1% for YOLOv6-s, 88.6% for YOLOv7, and 89.8% for YOLOv8.

The improved model therefore provided the highest mAP among the compared models. Its recall of 95.2% is particularly important for agricultural applications because missed disease cases can allow infection to spread. A high recall rate increases the probability that diseased leaves will be identified.

3.5 Generalization on PlantDoc and FieldPlant

The model was additionally evaluated using PlantDoc and FieldPlant to examine generalization. On PlantDoc, the proposed model achieved 69.0% precision, 83.4% recall, 75.5% F1-score, and 77.3% mAP. Its mAP was higher than YOLOX, YOLOv5, YOLOv6, YOLOv6-s, YOLOv7, and YOLOv8 by 7.7%, 11.8%, 3.4%, 5.7%, 4.3%, and 2.6%, respectively.

On FieldPlant, however, the proposed model achieved 66.5% mAP, only slightly lower than the 67.1% obtained by YOLOv6. The authors associate this difference partly with background information included in annotation boxes and possible noise introduced by cross-level feature fusion.

This finding is important because high performance on one dataset does not necessarily guarantee equivalent performance under all field conditions. Dataset diversity, image quality, annotation strategy, plant growth stage, geographic region, and environmental conditions can influence model performance.

4. Limitations and Future Scope

Although the attention and multi-scale fusion approach significantly improves tomato disease detection, several limitations remain. First, small lesions with similar visual characteristics, especially early blight and septoria leaf spot, remain challenging. Second, feature fusion may introduce redundant information or noise, potentially reducing precision in some cases. Third, datasets may not fully represent the diversity of agricultural environments encountered in different regions and seasons.

Future research should therefore expand datasets using images collected during different growth stages, seasons, geographical locations, and environmental conditions. Improved annotation strategies, including rotated bounding boxes, may also better represent the shape and orientation of leaves. Further optimization of feature-fusion modules could reduce noise and computational requirements. The incorporation of non-visual parameters such as temperature and humidity could additionally provide multimodal information for disease recognition. These directions are consistent with the future research recommendations reported in the source study.

5. Conclusion

The present study demonstrates the potential of combining attention mechanisms and multi-scale feature fusion for automated tomato leaf disease detection in complex natural environments. The integration of CBAM into YOLOv6 improves the network's ability to focus on disease-related spatial and channel features while reducing the influence of complex backgrounds. The proposed BiRepGFPN module complements this improvement by integrating shallow spatial information with deeper semantic information and thereby enhancing the detection of small lesions.

The complete improved YOLOv6 model achieved 92.9% precision, 95.2% recall, 94.0% F1-score, and 93.8% mAP on the tomato disease dataset. Its performance exceeded the baseline YOLOv6 model by 2.3 percentage points in precision, 4.0 percentage points in recall, 3.1 percentage points in F1-score, and 2.7 percentage points in mAP. The model also demonstrated improved generalization on PlantDoc.

Overall, attention-based feature selection and multi-scale feature fusion represent effective strategies for overcoming two major challenges in field-based plant disease detection: environmental interference and small disease lesions. Such approaches can contribute to the development of intelligent agricultural monitoring systems capable of supporting rapid and reliable disease diagnosis. Nevertheless, further work is required to improve robustness across different crops, geographical regions, environmental conditions, disease stages, and image-acquisition systems.

References

- Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2021). A novel deep learning method for detection and classification of plant diseases. *Complex & Intelligent Systems*, 8, 507–524. <https://doi.org/10.1007/s40747-021-00536-1>
- Cai, H., & Jiang, J. (2023). An improved plant disease detection method based on YOLOv5. In *2023 15th International Conference on Intelligent Human-Machine Systems and Cybernetics* (pp. 237–240). IEEE. <https://doi.org/10.1109/IHMSC58761.2023.00062>
- Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., & Sun, J. (2021). RepVGG: Making VGG-style ConvNets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13733–13742). IEEE. <https://doi.org/10.1109/CVPR46437.2021.01352>
- Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. *arXiv*. <https://doi.org/10.48550/arXiv.2107.08430>

- Guo, W., Feng, Q., Li, X., Yang, S., & Yang, J. (2022). Grape leaf disease detection based on attention mechanisms. *International Journal of Agricultural and Biological Engineering*, 15, 205–212. <https://doi.org/10.25165/j.ijabe.20221505.7548>
- Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13708–13717). IEEE. <https://doi.org/10.1109/CVPR46437.2021.01350>
- Jing, J., Li, S., Qiao, C., Li, K., Zhu, X., & Zhang, L. (2023). A tomato disease identification method based on leaf image automatic labeling algorithm and improved YOLOv5 model. *Journal of the Science of Food and Agriculture*, 103, 7070–7082. <https://doi.org/10.1002/jsfa.12793>
- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., et al. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv*. <https://doi.org/10.48550/arXiv.2209.02976>
- Li, S., Li, K., Qiao, Y., & Zhang, L. (2022). A multi-scale cucumber disease detection method in natural scenes based on YOLOv5. *Computers and Electronics in Agriculture*, 202, 107363. <https://doi.org/10.1016/j.compag.2022.107363>
- Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: A review. *Plant Methods*, 17, 1–18. <https://doi.org/10.1186/s13007-021-00722-9>
- Liu, J., & Wang, X. (2023). Tomato disease object detection method combining prior knowledge attention mechanism and multiscale features. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1255119>
- Qi, J., Liu, X., Liu, K., Xu, F., Guo, H., Tian, X., et al. (2022). An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Computers and Electronics in Agriculture*, 194, 106780. <https://doi.org/10.1016/j.compag.2022.106780>
- Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD* (pp. 249–253). ACM. <https://doi.org/10.1145/3371158.3371196>
- Wang, Y., Zhang, P., & Tian, S. (2024). Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion. *Frontiers in Plant Science*, 15, 1382802. <https://doi.org/10.3389/fpls.2024.1382802>