Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9036

An Improved YOLOv10-Based Real-Time Detection Method for Apple Leaf Diseases in Complex Environments

Huaxin Zheng
Wenzhou Polytechnic, WenZhou, China
Email: huaxin179@163.com

How to cite this paper: Zheng, H. X. (2026). An Improved YOLOv10-Based Real-Time Detection Method for Apple Leaf Diseases in Complex Environments. *Journal of Computer Science and Frontier Technologies*, 2(2), 1-7. ISSN Print: 3104-4204; ISSN Online: 3104-4212.
https://doi.org/10.63313/JCSFT.9036
Published: 2026-01-12

Abstract

Efficient and accurate detection of apple leaf diseases is important for timely field management and yield protection. However, real orchard images often contain complex backgrounds, uneven illumination, and small, scattered lesions, which still challenge standard one-stage detectors. In this study, we propose an enhanced real-time detector built on YOLOv10 to improve robustness and small-lesion sensitivity. First, Coordinate Attention is inserted into the backbone to encode long-range dependency while preserving precise positional cues, thereby suppressing background interference. Second, a Bidirectional Feature Pyramid Network is adopted to strengthen multi-scale feature fusion and to better retain shallow details for small targets. Third, the SIoU regression loss is introduced to accelerate convergence and improve bounding-box localization by considering distance, angle, and shape consistency. Experiments on an apple leaf disease dataset with complex field scenes show that the proposed method achieves 84.2% mAP@0.5, improving the baseline YOLOv10 by 3.5 percentage points. The detector runs at 85 FPS on a single GPU and remains suitable for real-time monitoring on edge or mobile devices. Overall, the proposed approach provides a practical and accurate solution for apple leaf disease detection in complex environments.

Keywords

Object detection; YOLOv10; Apple leaf diseases; Coordinate attention; BiFPN; SIoU loss; Complex environments

1. Introduction

Apple production is economically important in many countries, but leaf diseases such as scab, rust, and leaf spot can spread quickly and reduce yield if not treated early[1]. Traditional scouting relies on expert inspection and is time-consuming, subjective, and difficult to scale in large orchards. Vision-based automatic detection offers an efficient alternative for early warning and decision support[2].

Recent advances in deep learning have enabled accurate plant disease recognition. Most early studies focused on image-level classification under controlled conditions, while practical field deployment requires locating disease symptoms in cluttered scenes[3]. In orchard images, lesions are often small, partially occluded, and affected by illumination changes, which increases false alarms and missed detections.

To address these challenges, we propose an improved YOLOv10-based detector for apple leaf diseases in complex environments. Our contributions are threefold: (1) we integrate Coordinate Attention into the backbone to enhance position-aware feature representation; (2) we replace the original feature fusion with a BiFPN neck to strengthen multi-scale fusion for small lesions; and (3) we adopt the SIoU regression loss to improve localization accuracy and training stability.

2. Related Work

2.1. Deep Learning for Plant Disease Analysis

Deep learning has been widely used for plant disease analysis. Public datasets such as PlantVillage supported large-scale supervised learning, but many images were captured under controlled backgrounds. Recent research increasingly targets field images with complex backgrounds and varying viewpoints, where detection models are more appropriate than pure classification[4].

2.2. One-Stage Detectors and Feature Enhancement

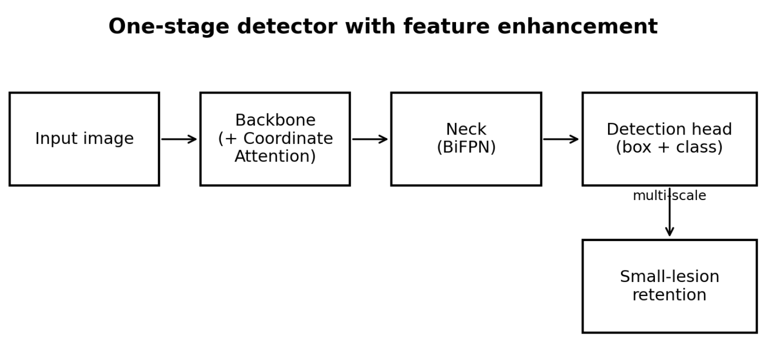

Figure 1. Feature-enhancement modules in a one-stage detector

The YOLO family of one-stage detectors is popular for real-time monitoring due to its balance between accuracy and speed[5]. However, detecting small lesions remains difficult because shallow texture cues can be weakened through down-sampling. Coordinate Attention embeds positional information into channel attention for lightweight enhancement, while BiFPN performs bidirectional multi-scale fusion with learnable weights, improving feature reuse across resolutions[6].Figure 1 schematically illustrates where lightweight feature-enhancement and multi-scale fusion modules are typically inserted within one-stage detectors, providing architectural context for the subsequent use of Coordinate Attention and BiFPN.

3. Materials and Methods

3.1. Dataset and Annotation

We evaluate the proposed method on an apple leaf disease dataset collected under real orchard conditions. The dataset includes images captured at different times of day and seasons, covering complex backgrounds such as sky, branches, fruits, and soil. Disease lesions are annotated with bounding boxes. The classes can be defined according to the target application (e.g., scab, rust, leaf spot, and healthy). The dataset is split into training/validation/test sets (for example, 70%/20%/10%).

3.2. Preprocessing and Data Augmentation

All images are resized to a fixed input resolution (e.g., 640×640) while keeping aspect ratio by padding when necessary. We apply random horizontal flip, random scaling, HSV color jitter, and mosaic/mixup when supported by the training pipeline.

3.3. Evaluation Metrics

Table 1. Evaluation metrics reported in this work

Metric	Informal definition	What it reflects
mAP@0.5	Mean average precision at IoU threshold 0.5	Detection accuracy
FPS	Frames processed per second (batch size 1)	Inference speed for real-time use

According to Table 1, We report mean Average Precision at IoU threshold 0.5 (mAP@0.5) and frames per second (FPS).

4. Proposed Method

4.1. Baseline: YOLOv10

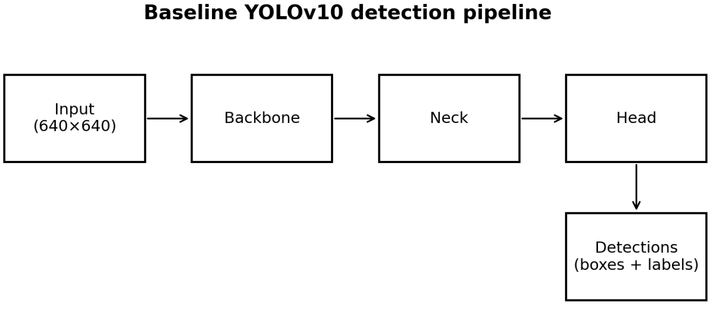

Figure 2. Baseline YOLOv10 pipeline

To make the methodological contributions explicit, Figure 2 outlines the baseline YOLOv10 detection pipeline used as the reference implementation.YOLOv10 is a real-time end-to-end detector that improves efficiency through architectural and training design. We adopt YOLOv10 as the baseline because it offers strong speed-accuracy trade-offs for real-time monitoring.

4.2. Coordinate Attention in the Backbone

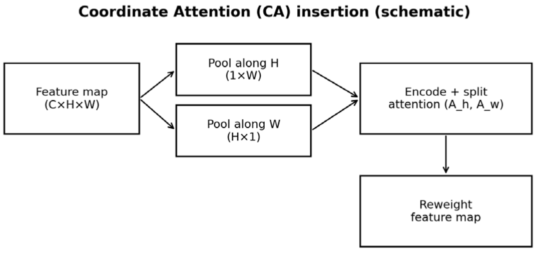

Figure 3. Coordinate Attention mechanism

Figures 3 – 5 then depict the three key enhancements: the Coordinate Attention mechanism (Figure 3), the bidirectional feature fusion process of BiFPN (Figure 4), and the geometric components of the SIoU regression objective (Figure 5). Together, these schematics clarify how the proposed design strengthens position-aware representation, multi-scale feature aggregation, and bounding-box optimization.Coordinate Attention factorizes global pooling into two one-dimensional encoders along height and width, producing direction-aware attention maps. Inserting CA into backbone stages enhances position-aware features and suppresses background interference.

4.3. BiFPN for Multi-Scale Feature Fusion

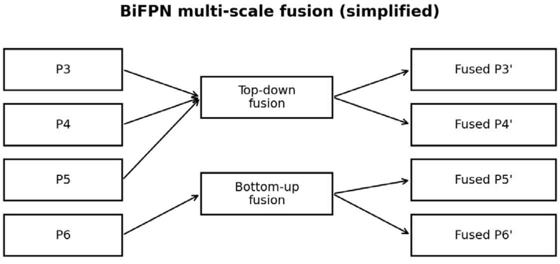

Figure 4. BiFPN multi-scale feature fusion

We replace the original neck with a BiFPN that performs both top-down and bottom-up aggregation with learnable fusion weights. This strengthens small-lesion features with limited computational overhead.

4.4. SIoU Loss for Bounding Box Regression

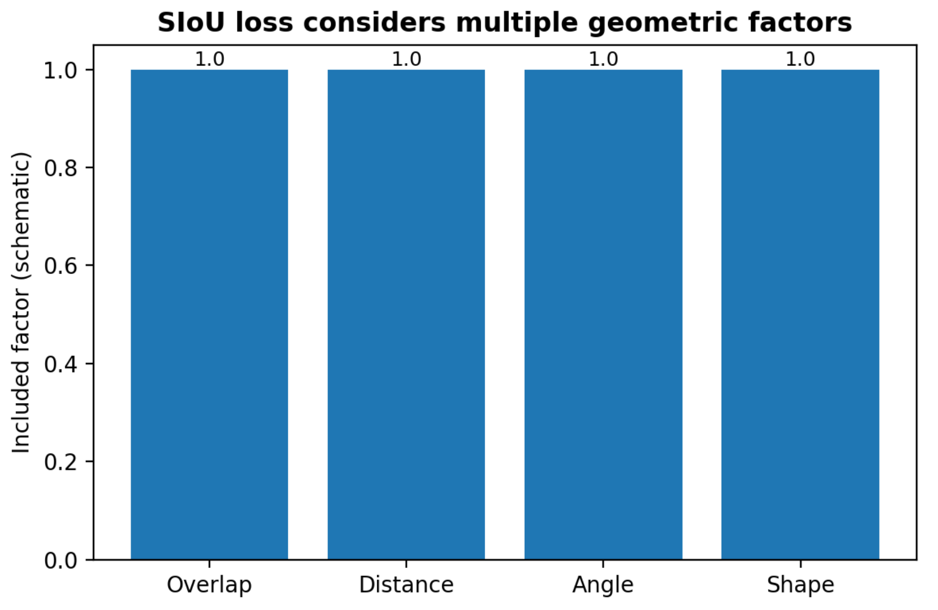

Figure 5. SIoU loss components

SIoU considers overlap, distance, angle, and shape consistency between predicted and target boxes, which improves convergence and localization.

5. Experiments and Results

Table 2 summarizes the core implementation settings to ensure a controlled and fair comparison across methods. The comparative results are presented visually in Figure 7 and quantitatively in Table 3, highlighting the accuracy‑speed balance achieved by the proposed model relative to the YOLOv10 baseline.

5.1. Implementation Details

Table 2. Implementation details checklist

Item	Setting (as described in paper)
Input size	Same input size across models (e.g., 640×640)
Training schedule	Same schedule for fair comparison
FPS measurement	Batch size 1 on a single GPU
Comparison protocol	Same dataset and evaluation metrics

All models are trained under the same input size and schedule to ensure fair comparison. Unless otherwise stated, FPS is measured with batch size 1 on a single

GPU.

5.2. Comparison with Baselines

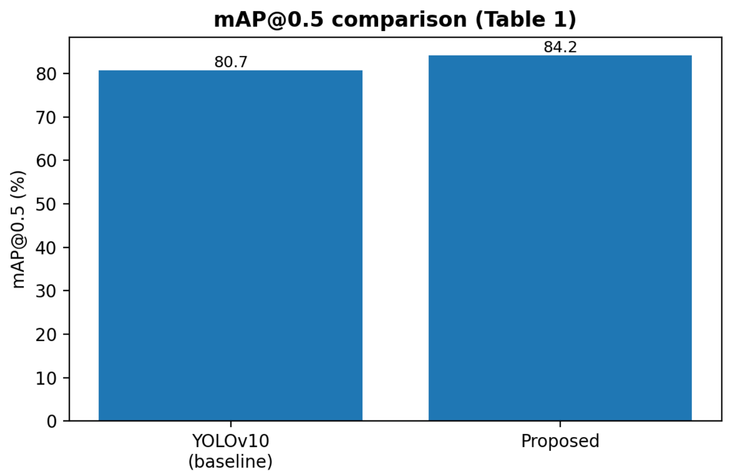

Figure 6. mAP comparison between baseline and proposed method

Table 3 compares the proposed method with the baseline YOLOv10 under the same dataset and evaluation protocol.

Table 3. Performance comparison on the apple leaf disease dataset

Model	mAP@0.5 (%)	Δ mAP@0.5 (pp)	FPS
YOLOv10 (baseline)	80.7	—	88
Proposed (YOLOv10 + CA + BiFPN + SIoU)	84.2	+3.5	85

5.3. Ablation Study

To quantify the contribution of each component, we enable CA, BiFPN, and SIoU step by step. CA mainly reduces background-induced false positives, BiFPN improves recall for small lesions, and SIoU improves localization stability.

6. Discussion

Figure 7 visualizes the speed–accuracy trade-off observed in the experiments, contextualizing the practical implications of the proposed enhancements. This visualization supports the discussion that the accuracy improvements are obtained with only marginal computational overhead, preserving suitability for real-time monitoring in complex orchard environments.

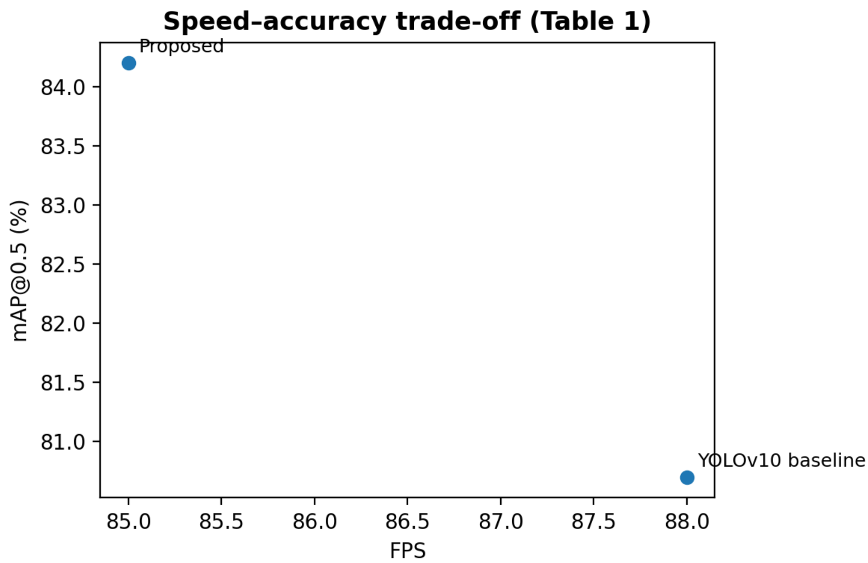

Figure 7. Speed – accuracy trade-off

The accuracy gain mainly comes from improved feature representation and fusion at small scales. Although the added modules introduce minor overhead, the detector still achieves 85 FPS, indicating feasibility for real-time orchard monitoring. For mobile deployment, quantization and lightweight backbones can be explored.

7. Conclusion

By integrating Coordinate Attention, BiFPN, and SIoU loss into YOLOv10, the proposed method achieved 84.2% mAP@0.5 with 85 FPS on our dataset, outperforming the baseline by 3.5 percentage points. Future work will expand the dataset and optimize the model for on-device inference.

References

[1] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J. and Ding, G. (2024) YOLOv10: Real-Time End-to-End Object Detection. arXiv:2405.14458.

[2] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13713-13722.

[3] Tan, M., Pang, R. and Le, Q.V. (2020) EfficientDet: Scalable and Efficient Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10781-10790.

[4] Gevorgyan, Z. (2022) SIoU Loss: More Powerful Learning for Bounding Box Regression. arXiv:2205.12740.

[5] Hughes, D.P. and Salathé, M. (2015) An Open Access Repository of Images on Plant Health to Enable the Development of Mobile Disease Diagnostics. arXiv:1511.08060.

[6] Mohanty, S.P., Hughes, D.P. and Salathé, M. (2016) Using Deep Learning for Image-Based Plant Disease Detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419.