International Core Journal of Engineering Volume 12 Issue 4, 2026
ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

Intelligent Detection of Northern Shaanxi Red Fuji Apples based on Improved YOLOv8 Model

Lina Zhang[a], Jiale Zhang[b], Yachen Zhao[c], Jing Hong[d], and Jingyuan He[e, *]

School of Mathematics and Computer Science, Yan'an University, Yan'an 716000, China

[a]amazingzhanglina@yau.edu.cn, [b]vector060227@gmail.com, [c]zhaochuji28766@gmail.com, [d]3601904526@qq.com, [e]18992118537@163.com

[*]Corresponding author: Jingyuan He

Abstract

Detection of Northern Shaanxi Red Fuji apples in natural orchards faces challenges such as branch-leaf occlusion, fruit overlapping, illumination variation, and scale differences, leading to decreased detection accuracy and localization stability of the model. To address these issues, YOLOv8s was employed as the baseline model, with improvements made from three aspects: small object detection layer, attention mechanism, and bounding box regression loss. Specifically, a P2 detection layer, Coordinate Attention, EIoU, and WIoU were introduced. Multiple ablation experiments were conducted under unified training conditions, and best.pt was adopted for re-validation to avoid deviations caused by early stopping. Experimental results demonstrate that among the improved schemes, YOLOv8s+EIoU achieved the optimal performance, with mAP50-95 increased to 69.73%, representing a 0.21 percentage point improvement over the baseline, which verifies the optimization effect of EIoU on bounding box regression. This study provides technical references for intelligent apple recognition, automated harvesting, and orchard management.

Keywords

Northern Shaanxi Red Fuji Apples; YOLOv8; Object Detection; EIoU; Smart Agriculture.

1. Introduction

Apple detection is one of the fundamental tasks in intelligent orchards. Its performance affects not only fruit counting and maturity analysis, but also the path planning of harvesting robots and the efficiency of fine-grained orchard management. Red Fuji apples are widely grown in Northern Shaanxi, and the corresponding orchard scenes are often large-scale and visually complex. For this reason, it is of clear practical importance to develop a detection method that works reliably under natural conditions.

Earlier studies on apple detection mainly relied on handcrafted color, shape, and texture features for target segmentation. These methods can work reasonably well under simple backgrounds or stable illumination, but in real orchards they are easily affected by shadows, occlusion, and overlapping fruit. With the rapid development of deep learning, convolutional neural network based detectors have gradually become the mainstream choice. Among them, the YOLO family has been widely used in fruit detection and harvesting-related tasks because of its fast inference speed, compact architecture, and convenient deployment [1-4]. Existing apple detection studies have mainly focused on lightweight architectures, multi-scale feature fusion, and robustness under complex backgrounds. For example, Yan et al. [1] investigated apple target recognition for robotic picking based on an improved YOLOv5 model; Zhang et al. [2] enhanced apple detection in complex backgrounds within the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

YOLOv4 framework; and Sun et al. [3] and Liu et al. [4] improved orchard apple recognition from the perspectives of modified YOLOv5s and YOLOv8 structures, respectively. In the YOLOv8 stage, related studies paid more attention to occlusion, small targets, and real-time deployment in complex orchard environments [5-8]. In addition, some work explored auxiliary-task learning to improve apple detection. Zhao et al. [9] introduced fallen apple detection as an auxiliary task during training and showed that task collaboration can improve robustness. As a relatively recent member of the YOLO family, YOLOv8 further improves feature extraction, feature fusion, and head design, and its overall performance is more balanced than that of earlier versions. Even so, several typical problems remain when YOLOv8 is applied directly to Northern Shaanxi Red Fuji apple detection. First, distant or densely distributed apples are small and may lose useful information during downsampling. Second, leaves and branches create occlusion and texture interference, which can lead to false detections. Third, in a single-class apple detection task, bounding-box regression quality strongly affects the final result, and once localization becomes inaccurate, mAP50-95 drops noticeably. Based on these observations, this study carries out a series of improvement experiments around YOLOv8s. The work first examines whether a P2 detection layer can strengthen small-target detection, then tests Coordinate Attention for better feature representation under complex backgrounds, and finally compares EIoU and WIoU from the perspective of bounding-box regression. The goal is to identify a scheme that is genuinely more suitable for Northern Shaanxi Red Fuji apple detection, rather than simply making the model more complex by stacking modules.

The main contributions of this paper are as follows: (1) several improved YOLOv8s schemes were designed for natural Red Fuji apple scenes characterized by small targets, occlusion, and localization deviation; (2) all comparisons were conducted using revalidated best.pt results, and the effects of each scheme on Precision, Recall, mAP50, and mAP50-95 were analyzed; and (3) based on the experimental results, a more suitable model direction for apple detection was identified and its stability was discussed.

2. Problem Analysis

The images of Northern Shaanxi Red Fuji apples possess prominent natural scene characteristics. The apple targets vary significantly in scale, with some small-sized objects; occlusion by branches and leaves as well as fruit overlapping occur frequently. Affected by shooting angles and weather conditions, the images also suffer from backlighting, shadows, and uneven brightness. If the model fails to handle such complex scenarios properly, two typical problems will arise: missed detection, or offset and inaccurate fitting of bounding boxes for detected targets.

The YOLO family is a typical one-stage object detection framework. Its core idea is to predict object categories and locations in a single forward pass. As a result, it can maintain relatively high detection accuracy while preserving fast inference speed [10]. YOLOv8 consists of three main parts: the Backbone, the Neck, and the Head. The Backbone extracts multi-level features from the input image, the Neck integrates information from different scales through upsampling, downsampling, and feature fusion, and the Head outputs category and bounding-box predictions. Its multi-scale fusion design is closely related to structures such as FPN and PAN [11-12]. Compared with earlier versions, YOLOv8 adopts anchor-free localization and a decoupled head for classification and regression, which helps balance detection accuracy and inference efficiency.

Therefore, according to the practical requirements of the apple detection task, this study selects YOLOv8s as the baseline model. Firstly, YOLOv8s has a moderate parameter scale with low costs for training and validation. Secondly, the baseline model already possesses competitive detection performance, making it suitable for analyzing the actual effectiveness of different improvement strategies. The specific improvements are as follows: (1) Introduce the P2 detection layer to supplement high-resolution shallow features, aiming to enhance the detection accuracy of small targets; (2) Adopt Coordinate Attention to strengthen the model's focus on target regions and suppress

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

background interference; (3) Utilize EIoU and WIoU to optimize the bounding box regression process, with emphasis on improving localization performance under strict IoU constraints.

3. Improved YOLOv8

3.1 P2 Detection Layer

In the original YOLOv8 architecture, detection relies mainly on feature maps at multiple scales for prediction. Deep feature maps possess stronger semantic representation capability but retain relatively limited details of small targets. When apples appear in long-distance or dense distributions, they usually have small sizes and unclear boundaries; direct use of the original structure easily leads to missed detections. This paper introduces the P2 detection layer, enabling the network to perform prediction on higher-resolution feature maps. It preserves more shallow edge and texture features and enhances the model's perception of small-sized apples. However, this scheme also has potential drawbacks: shallow features contain more background interference. Without proper suppression, the model tends to generate redundant bounding boxes.

To avoid presenting the P2 layer as merely an extra small-object branch, the detection-head structure shown in Fig. 1. Compared with the original YOLOv8s detector, which predicts on P3, P4, and P5, this structure introduces a higher-resolution P2/4 output, expanding the final Detect stage from three scales to four. This is consistent with the experimental setting used here for small-apple detection.

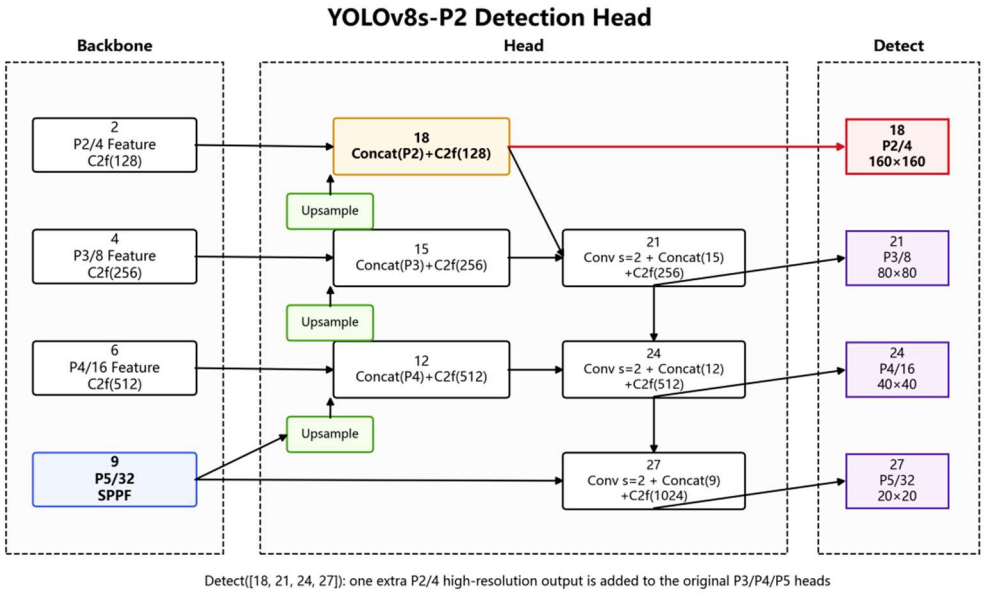

Fig. 1 Structure of the YOLOv8s-P2 detection head

3.2 Coordinate Attention Mechanism

Natural orchard backgrounds are complex. Branches, leaves, and fruits often overlap locally, and convolutional features are easily disturbed by background textures during extraction. Coordinate Attention introduces positional information on top of channel attention [13], so it can preserve a degree of coordinate awareness along two spatial directions. For this reason, it is considered more suitable for visual tasks in which object location matters. This paper introduces Coordinate Attention into both the original YOLOv8s network and the added P2 branch. The purpose is to enable the model to more accurately distinguish apple targets from surrounding background while maintaining satisfactory feature representation capability. Coordinate Attention is not adopted to increase network complexity; instead, it achieves targeted feature recalibration, thereby reducing false detections and alleviating interference caused by occlusion.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

This study does not insert attention blocks throughout the entire YOLOv8s network. Instead, following the actual configuration in `custom_models/yolov8s-ca.yaml`, `CoordAtt` is inserted only before the three output scales in the Head. The left side of Fig. 2 shows the actual insertion points used in the experiments, while the right side illustrates the core mechanism of Coordinate Attention. In this way, the model retains channel information while also introducing direction-aware positional information, which strengthens feature representation for apple target regions.

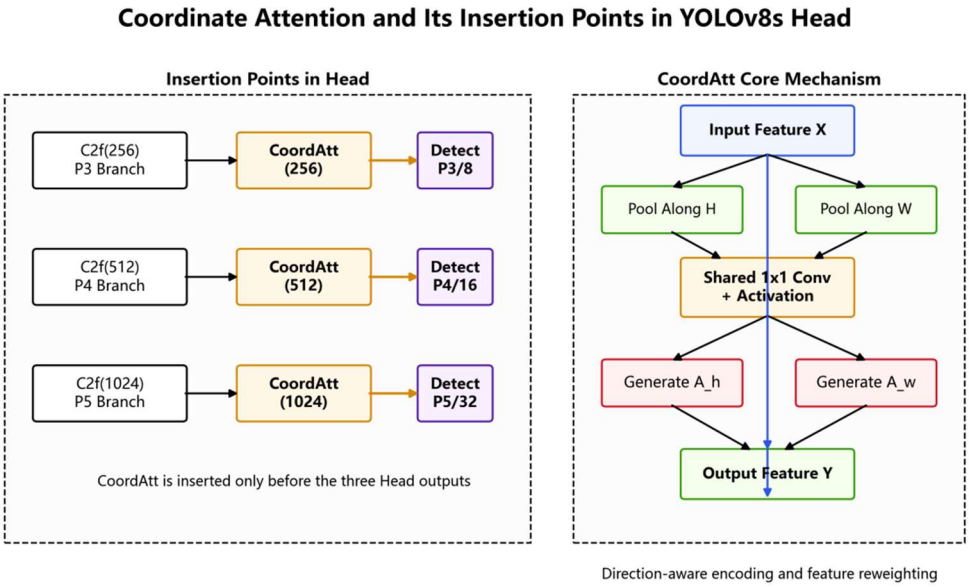

Fig. 2 Coordinate Attention and its insertion points in the YOLOv8s head

3.3 Regression Loss Optimization

In apple detection, the quality of bounding-box regression has a strong influence on the final metrics. This is especially true for mAP50-95, which does not only ask whether an apple is found, but also whether the predicted box fits it accurately. YOLOv8 uses CIoU as the default regression loss. Although CIoU is stable in general detection tasks, width-height deviation and center offset can still appear when apple boundaries are crowded or local occlusion is obvious.

EIoU includes width and height errors explicitly in the optimization process, in addition to IoU and center-distance constraints [14]. From the task perspective, this is better aligned with apple detection, because even though apples belong to a single category, the requirement for box regression accuracy is still high. In addition to EIoU, WIoU [15] was also tested to examine the suitability of dynamic focusing regression losses on the current dataset. The EIoU regression loss can be written as follows:

$$L\_EIoU = 1 - IoU + \rho^2(b,b^{gt})/c^2 + (w - w^{gt})^2/c\_w^2 + (h - h^{gt})^2/c\_h^2 \quad (1)$$

In Eq. (1), L_EIoU denotes the EIoU bounding-box regression loss; IoU denotes the intersection-over-union between the predicted box and the ground-truth box; $\rho^2(b,b^{gt})$ denotes the squared Euclidean distance between the predicted box center b and the ground-truth box center b^gt; c denotes the diagonal length of the minimum enclosing rectangle covering both boxes; w and h denote the width and height of the predicted box; w^gt and h^gt denote the width and height of the ground-truth box; and c_w and c_h denote the width and height of the minimum enclosing rectangle, respectively.

To clarify the loss modification, Fig. 3 illustrates the regression path comparison between CIoU and EIoU according to the actual training procedure. It should be noted that `exp6` does not alter the network structure of YOLOv8s. Instead, it replaces the default CIoU with EIoU inside `BboxLoss`,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

so that bounding-box regression explicitly considers width and height errors in addition to IoU and center-distance constraints. This change is consistent with the fact that apple detection in this study is highly sensitive to bounding-box fit.

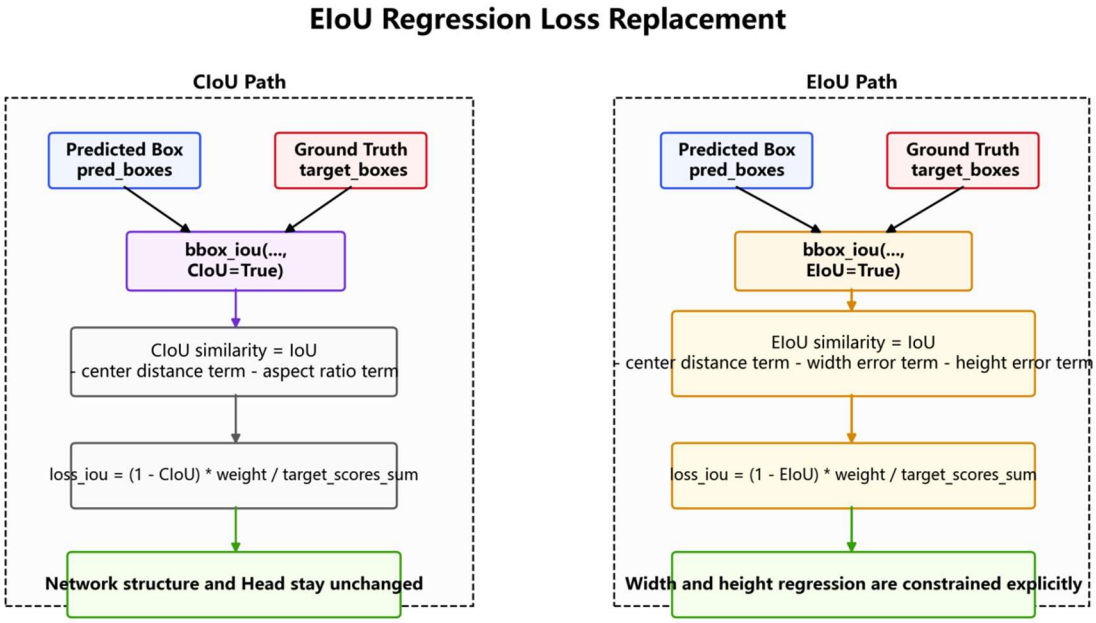

Fig. 3 EIoU regression loss replacement pipeline

4. Experiments and Result Analysis

4.1 Dataset and Evaluation Metrics

Fig. 4 Sample images from the Northern Shaanxi Red Fuji apple dataset

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

The experiments in this study use a Northern Shaanxi Red Fuji apple image dataset, whose annotations have been completed and converted into the YOLO training format. The dataset is split into 3,244 training images, 316 validation images, and 166 test images. The samples cover multiple situations, including unobstructed fruit, partial occlusion, overlapping apples, and different lighting conditions, and therefore reflect the real difficulties of apple detection in natural orchards reasonably well.

To show the composition of the dataset more intuitively, several representative images were selected from the validation set and displayed together with manual bounding-box annotations, as shown in Fig. 4. The examples include not only scenes with a small number of targets and relatively simple backgrounds, but also scenes with densely distributed fruit, local occlusion, and obvious scale differences. These samples reflect the complexity of natural orchard images of Northern Shaanxi Red Fuji apples quite well.

The evaluation metrics used in this paper are Precision, Recall, mAP50, and mAP50-95. Precision mainly reflects the proportion of correct detections in the predicted results, Recall reflects the ability of the model to find apple targets, mAP50 evaluates the mean precision at an IoU threshold of 0.5, and mAP50-95 provides a stricter and more comprehensive measure of overall localization performance across multiple IoU thresholds.

4.2 Experimental Environment and Parameter Settings

All experiments were carried out under Windows 11. The GPU was an NVIDIA GeForce RTX 4060 Laptop GPU, the deep learning framework was PyTorch, and the training and validation pipeline was implemented with Ultralytics YOLOv8. Unless otherwise stated, all experimental groups used the same settings: input size 640×640, batch size 8, 100 epochs, AdamW as the optimizer, an initial learning rate of 0.0005, and an EarlyStopping patience value of 40.

A further point should be clarified here. Some runs triggered EarlyStopping, so the last row in 'results.csv' does not always match the actual performance of the saved 'best.pt' model. To ensure a fair comparison, all completed experiments were revalidated with `best.pt`, and every table and textual analysis in this paper is based on those revalidated results.

4.3 Ablation Study

Table 1. Ablation study results

Group	Model variant	Precision/%	Recall/%	mAP50/%	mAP50-95/%
1	YOLOv8s	89.08	88.37	93.27	69.52
2	YOLOv8s+P2	88.96	86.96	93.13	69.11
3	YOLOv8s+P2+ imgsz800	89.08	89.21	94.08	68.49
4	YOLOv8s+P2+CA	89.12	87.53	92.80	68.85
5	YOLOv8s+P2+CA + EIoU	89.03	88.02	93.74	68.71
6	YOLOv8s + CA	90.48	85.84	92.79	68.75
7	YOLOv8s + EIoU	88.91	87.60	92.95	69.73
8	YOLOv8s + WIoU	89.28	88.40	93.55	69.24

As shown in Table 1, the YOLOv8s baseline reaches an mAP50-95 of 69.52%, which serves as the reference point for the following schemes. After the P2 detection layer is added in Group 2, mAP50-95 becomes 69.11%, which does not exceed the baseline. This suggests that, for the current apple dataset, simply adding a shallow detection branch does not directly translate into better overall localization performance.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

In Group 3, the input size is increased to 800 on top of P2. Recall rises to 89.21% and mAP50 rises to 94.08%, which shows that the model does become better at finding more apple targets. However, mAP50-95 drops to 68.49%. This indicates that more detections do not necessarily mean better localization. The gain from higher resolution mainly appears in recall rather than in detection quality under strict IoU thresholds.

Groups 4 and 5 introduce Coordinate Attention and EIoU into the P2 branch, respectively. The results show that P2+CA reaches an mAP50-95 of 68.85%, while P2+CA+EIoU reaches 68.71%. Both are slightly better than P2+800, but still do not exceed the baseline. This suggests that under the P2 route, the overall gain remains limited even when attention or regression-loss optimization is added. In other words, the key issue may lie less in adding another branch than in making box regression more stable.

Group 6 adds Coordinate Attention only to the baseline model. Its Precision reaches 90.48%, which indicates that CA helps reduce false detections to some extent, but Recall drops to 85.84% and mAP50-95 is only 68.75%. This result suggests that the attention mechanism is not ineffective in this task, but the gain it provides is still not enough to push the overall metric beyond the baseline.

After EIoU is introduced into YOLOv8s in Group 7, mAP50-95 reaches 69.73%, the highest value among all completed improvement schemes and 0.21 percentage points above the baseline. This is the most noteworthy result of the present experiments. As shown in Eq. (1), EIoU penalizes width and height errors in addition to IoU and center-distance constraints. Compared with structural changes such as P2 and CA, it acts more directly on bounding-box regression and better matches the practical requirement for localization precision in apple detection.

After WIoU is adopted in Group 8, mAP50-95 becomes 69.24%. Although this is higher than most of the P2-based and CA-based schemes, it is still lower than EIoU. This means that, on the present dataset, WIoU does not show a clearer advantage than EIoU. As for Soft-NMS, it led to an obvious drop in the overall metrics under the current settings, so it was not pursued further as a main improvement direction.

4.4 Repeated Experiment Analysis

Table 2. Results of repeated experiments

Model	Run	Precision/%	Recall/%	mAP50/%	mAP50-95/%
YOLOv8s	Run 1	89.08	88.37	93.27	69.52
YOLOv8s	Run 2	89.85	86.51	92.94	69.54
YOLOv8s + EIoU	Run 1	88.91	87.60	92.95	69.73
YOLOv8s + EIoU	Run 2	90.23	88.62	93.87	69.35

To avoid drawing conclusions from a single best run alone, repeated experiments were further carried out for the YOLOv8s baseline and the YOLOv8s+EIoU model. As shown in Table 2, the baseline model reaches mAP50-95 values of 69.52% and 69.54% in two runs, which are very close. This indicates that YOLOv8s is stable on the current dataset.

By contrast, YOLOv8s+EIoU reaches 69.73% mAP50-95 in the first run, but 69.35% in the second. In other words, EIoU does produce the highest result in the best single run, but it does not maintain the same margin in repeated experiments. This suggests that EIoU is better viewed as a valid optimization direction for the current task, rather than as a fully stable final answer that can replace the baseline without reservation. From a research perspective, this result is not contradictory. The baseline shows that YOLOv8s itself is already strong, while EIoU indicates that there is still room for improvement through regression-loss optimization without making large structural changes. In contrast, P2 and Coordinate Attention each provide local benefits, but neither shows the same advantage in the overall metric.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

4.5 Detection Result Analysis

To present the practical behavior of the models more directly in the apple detection task, several typical images from the test set were selected for visual analysis. The detection results of the YOLOv8s baseline and the improved YOLOv8s+EIoU model are shown in Fig. 5.

(a) Detection results of the YOLOv8s baseline on the test set

(b) Detection results of the YOLOv8s+EIoU model on the test set

Fig. 5 Comparison of detection results between YOLOv8s and YOLOv8s+EIoU

As shown in Fig. 5(a), the YOLOv8s baseline can already recognize and localize apples quite accurately in scenes with clear boundaries and weak occlusion. However, missed detections and box offsets still appear under local occlusion, small-target conditions, and scenes where fruits are close to each other. As shown in Fig. 5(b), after the EIoU loss is introduced, the predicted boxes fit the targets better in some complex scenes. The improvement is especially visible when fruit contours are less clear or targets are close to each other, where the predicted boxes match the apple boundaries more steadily. This indicates that EIoU provides a useful optimization effect on the bounding-box regression process.

The comparison in Fig. 5 further shows that the improved model is not clearly better than the baseline on every image. Even so, in several representative samples, its predicted boxes are closer to the true target extent and overlap better with the actual fruit region. This is consistent with the quantitative results reported earlier, where the EIoU scheme achieved the highest single-run mAP50-95.

Taken together with Tables 1 and 2, the results show that different improvement strategies address different aspects of the task. P2 and higher-resolution input tend to improve the detection rate of apple targets, especially when the targets are small or densely distributed, and this route more easily increases Recall and mAP50. At the same time, however, localization error and redundant boxes also increase, which eventually leads to a lower mAP50-95.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

The role of Coordinate Attention is mainly reflected in feature representation. Judging from the experimental results, the CA-based schemes perform relatively well in Precision, which suggests that CA helps suppress background interference and reduce false detections. However, for a single-class target such as apples, whether the model can exceed the baseline depends more on the accuracy of box regression than on attention enhancement alone. EIoU shows a more direct match to the current task. Although apple targets belong to a single class, width-height deviation and center deviation under occlusion and overlap conditions directly affect detection quality. As can be seen from Eq. (1), EIoU constrains the width term and the height term separately, so the model can optimize size differences more directly during bounding-box regression. This also helps explain why EIoU achieves the highest mAP50-95 in the best single run.

Overall, the full set of experiments suggests that continuing to stack additional modules on the P2 route does not provide an obvious return for Northern Shaanxi Red Fuji apple detection. A more effective strategy is to keep the overall YOLOv8s structure stable while making targeted improvements to the bounding-box regression process.

5. Conclusion

To address missed detection of small targets, background interference, and localization deviation of Northern Shaanxi Red Fuji apples in natural orchards, this paper takes YOLOv8s as the baseline model and carries out a systematic study around the P2 detection layer, Coordinate Attention, and bounding-box regression loss optimization. The main conclusions are as follows:

(1) The YOLOv8s baseline already shows good stability on the current apple dataset. The two repeated runs reach mAP50-95 values of 69.52% and 69.54%, which indicates that it has strong repeatability as a reference model.

(2) The P2 detection layer and higher-resolution input can improve the ability to find apple targets to some extent, but these methods mainly raise Recall and mAP50, while their contribution to overall localization performance under strict IoU conditions is limited.

(3) Coordinate Attention does improve feature representation under complex backgrounds to some extent, but whether used alone or together with P2, its overall metrics do not exceed the baseline model. This suggests that the gain provided by this mechanism is relatively limited in the current task.

(4) Among all completed improvement schemes, YOLOv8s+EIoU achieves the highest single-run mAP50-95 of 69.73%, which is 0.21 percentage points higher than the baseline model, indicating that EIoU can improve the quality of apple bounding-box regression to a certain extent. However, the repeated experiments also show that this scheme still has some fluctuation, so further optimization is needed in terms of training stability and task adaptation.

Overall, the EIoU-optimized YOLOv8 model shows practical potential for the intelligent detection of Northern Shaanxi Red Fuji apples. Future work may continue along several directions, including more targeted data augmentation, scene-wise error analysis, and inference-stage optimization, in order to further improve model stability and overall performance.

Acknowledgments

Project Funding: Supported by the 2025 Undergraduate Innovation and Entrepreneurship Training Program of Yan'an University (D2025123).

References

[1] B. Yan, P. Fan, X. Lei, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5. Remote Sensing. Vol. 13 (2021) No. 9, p. 1619.

[2] C. Zhang, F. Kang, Y. Wang. An improved apple object detection method based on lightweight YOLOv4 in complex backgrounds. Remote Sensing. Vol. 14 (2022) No. 17, p. 4150.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0038

[3] X. Sun, Y. Zheng, D. Wu, et al. Detection of orchard apples using improved YOLOv5s-GBR model. Agronomy. Vol. 14 (2024) No. 4, p. 682.

[4] Z. Liu, R.M.R.D. Abeyrathna, R.M. Sampurno, et al. Faster-YOLO-AP: A lightweight apple detection algorithm based on improved YOLOv8 with a new efficient PDWConv in orchard. Computers and Electronics in Agriculture. Vol. 223 (2024), p. 109118.

[5] [B. Yan, Y. Liu, W. Yan. A novel fusion perception algorithm of tree branch/trunk and apple for harvesting robot based on improved YOLOv8s. Agronomy. Vol. 14 (2024) No. 9, p. 1895.

[6] T. Wu, Z. Miao, W. Huang, et al. SGW-YOLOv8n: An improved YOLOv8n-based model for apple detection and segmentation in complex orchard environments. Agriculture. Vol. 14 (2024) No. 11, p. 1958.

[7] M. Wang, F. Li. Real-time accurate apple detection based on improved YOLOv8n in complex natural environments. Plants. Vol. 14 (2025) No. 3, p. 365.

[8] H. Wu, X. Mo, S. Wen, et al. DNE-YOLO: A method for apple fruit detection in diverse natural environments. Journal of King Saud University-Computer and Information Sciences. Vol. 36 (2024), p. 102220.

[9] J. Zhao, A. Lipani, C. Schillaci. Fallen apple detection as an auxiliary task: Boosting robotic apple detection performance through multi-task learning. Smart Agricultural Technology. Vol. 8 (2024), p. 100436.

[10] J. Redmon, A. Farhadi. YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767. (2018).

[11] T.Y. Lin, P. Dollar, R. Girshick, et al. Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017), p. 2117-2125.

[12] S. Liu, L. Qi, H. Qin, et al. Path aggregation network for instance segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2018), p. 8759-8768.

[13] Q. Hou, D. Zhou, J. Feng. Coordinate attention for efficient mobile network design. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2021), p. 13713-13722.

[14] Y.F. Zhang, W. Ren, Z. Zhang, et al. Focal and efficient IoU loss for accurate bounding box regression. Neurocomputing. Vol. 506 (2022), p. 146-157.

[15] Z. Tong, Y. Chen, Z. Xu, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. arXiv preprint arXiv:2301.10051. (2023).