CITRUS FLOWER, FRUIT, AND SHOOT RECOGNITION BASED ON IMPROVED YOLOv10
/
基于改进 YOLOv10 的柑橘花果梢识别

Wenfeng GUO*[1], ZhiFang BI[1], Linjuan WANG[1], Qinqin WU[2], Han WANG[3]
[1] Department of Basic Science, Shanxi Agricultural University, Taigu, Shanxi / China;
[2] School of Mega Data, Jiangxi Institute of Fashion Technology, Nanchang, Jiangxi / China;
[3] College of Information Science and Engineering, Shanxi Agricultural University, Taigu, Shanxi / China;
Corresponding Authors: Wenfeng Guo; E-mail: wenfengguo@sxau.edu.cn
DOI: https://doi.org/10.35633/inmateh-79-47

Keywords: *YOLOv10; deep learning; Citrus; BAM; GIoU*

ABSTRACT
In the process of agricultural intelligence, precise detection of plant organs serves as the foundation for core tasks such as crop phenotyping analysis and yield prediction. However, in complex field environments, small targets such as citrus flowers and shoots face challenges including scale variation, background interference, and dense occlusion, which severely impact detection accuracy. This study improves the YOLOv10 model by introducing the BAM (Bottleneck Attention Module) attention mechanism and GIoU (Generalized Intersection over Union) loss function, constructing a YOLOv10s-BAM-GIoU model suitable for citrus flower, fruit, and shoot recognition. The BAM attention mechanism enhances the model's feature extraction capability for small target organs under complex backgrounds through parallel channel and spatial attention branches; the GIoU loss function improves the localization accuracy of densely occluded targets by optimizing the geometric alignment between predicted and ground-truth boxes. Validation experiments were conducted on a self-constructed dataset. The experimental results show that the improved YOLOv10s achieves significant advantages in comprehensive detection accuracy, with an mAP50 of 89.1%, representing an improvement of 2.9%~9.5% over the original YOLOv10s and other comparative models. In fine-grained category detection, the model achieves mAP50 of 91.2%, 83.6%, and 92.5% for shoots, flowers, and fruits, respectively. Furthermore, while maintaining high detection accuracy, the model achieves a detection speed of 23.6 ms per frame, meeting real-time detection requirements. The research results demonstrate that the improved YOLOv10s model integrating the BAM attention mechanism and GIoU loss function achieves an optimal balance between accuracy and speed in citrus organ detection tasks, providing a preferred solution for field real-time detection systems.

摘要
在农业智能化进程中，植物器官精准检测是作物表型分析、产量预测等核心任务的基础。然而，田间复杂环境下柑橘花、梢等小目标存在尺度多变、背景干扰及密集遮挡等问题，严重影响检测精度。本研究基于 YOLOv10 模型，通过引入 BAM 注意力机制和 GIoU 损失函数对其进行改进，构建了适用于柑橘花果梢识别的 YOLOv10s-BAM-GIoU 模型。其中，BAM 注意力机制通过通道与空间的并行注意力分支，增强模型对复杂背景下小目标器官的特征提取能力；GIoU 损失函数则通过优化预测框与真实框的几何对齐，提升模型对密集遮挡目标的定位精度。在本研究自建数据集上进行验证，实验结果显示，改进后的 YOLOv10s 在综合检测精度上展现显著优势，其 mAP50 达到 89.1%，较原始 YOLOv10s 及其他对比模型提升 2.9%~9.5%。在细分类别检测中，该模型对梢（shoot）、花（flower）、果实（fruit）的 mAP50 分别达到 91.2%、83.6%、92.5%。此外，模型在保持较高检测精度的同时，检测速度达到 23.6ms/帧，满足实时检测需求。研究结果表明，融合 BAM 注意力机制与 GIoU 损失函数的改进 YOLOv10s 模型在柑橘器官检测任务中实现了精度与速度的最佳平衡，为田间实时检测系统提供了优选方案。

INTRODUCTION
The citrus industry is a vital component of agricultural economy. However, traditional cultivation management relies on manual experience, making it difficult to precisely regulate the proportion of flowers and fruits, leading to significant yield fluctuations and inconsistent fruit quality. Accurate identification of the number of flowers, fruits, and shoots is a prerequisite for implementing precision management measures such as flower and fruit thinning and scientific pruning.

Nevertheless, citrus flowers, fruits, and shoots share similar morphology and color and are often distributed in clusters. Traditional manual counting methods are inefficient and highly subjective, failing to meet the refined requirements of modern smart agriculture. The precise recognition of citrus flowers, fruits, and shoots enables scientific regulation of flower and fruit loads, thereby enhancing citrus quality and promoting the digital transformation of the industry.

Addressing the challenges of refined recognition that traditional methods struggle to solve, researchers have conducted extensive studies based on deep learning technologies in recent years, achieving significant progress. Huang et al. proposed ORD-YOLO, an enhanced YOLOv8-based model for citrus maturity detection. The key innovations of their approach include the incorporation of ODConv for robust feature extraction, RepGFPN for multi-scale feature fusion, and a Dynamic Head for adaptive key feature perception (*Huang et al., 2025*).

Lin et al. enhanced YOLOv5s by incorporating CBAM and Contextual Transformer modules into its backbone network and substituting GIoU with SIoU as the loss function. These modifications improved the average recognition accuracy for citrus fruits by 3.51% compared to the original algorithm (*Lin et al., 2023*).

For accurate assessment of citrus fruit ripeness (unripe and fully ripe), *Ünal et al.* proposed a lightweight deep learning model featuring an optimized YOLO12n-Seg backbone. By embedding a GhostConv module into the neck to minimize parameter overhead and applying a Global Attention Mechanism (GAM) to enhance feature discrimination, the model effectively reduces background clutter while maintaining high detection fidelity (*Ünal and Eceoğlu, 2025*). To accurately and immediately recognize mature citrus, *Xu et al.* proposed a novel method based on the Otsu adaptive threshold method and on the improved random ring method (*Xu et al., 2020*).

Lyu et al. achieved precise detection of green citrus fruits in natural scenes by proposing the YOLO-GC model incorporating GhostNet and GAM attention mechanisms, achieving a model mAP of 97.8%, yet missed detections occurred in densely overlapped fruit areas (*Lyu et al., 2024*).

El Akrouchi et al. broke whole-tree high-resolution images of small green citrus fruits into smaller patches to capture finer details, and employed a model combining Multiscale Vision Transformer version 2 (MViTv2) with Cascade Mask R-CNN for fruit detection (*El Akrouchi et al., 2025*).

Xiao et al. achieved high-precision citrus recognition in orchard environments by proposing the AC-YOLO model with a self-attention mechanism, attaining an average precision of 97.27%, though recognition performance for severely occluded fruits remained suboptimal (*Xiao et al., 2024*). A yield prediction scheme was proposed by *Qin et al*., which employs both Faster R-CNN and SSD to detect citrus fruits on the tree and subsequently computes fruit size based on the short diameter measurement (*Qin et al., 2021*).

Subeesh et al. used unmanned aerial vehicles (UAVs) to capture images of citrus fruits, and employed multiple deep learning algorithms such as YOLOv7 to detect two categories of fruits, "harvest-ready" and "unripe," from canopy images, thereby enabling orchard yield estimation (*Subeesh et al., 2024*). Based on the mask regional convolutional neural network (Mask R-CNN) and a branch segment merging algorithm, an integrated system was developed to simultaneously detect and measure citrus fruits and branches (*Yang et al., 2020*).

Qiu et al. achieved automatic segmentation and phenotypic quantification of citrus floral organs by independently developing the CF-ASPM model, enabling fine segmentation of seven organ types including petals, styles, and ovaries (*Qiu et al., 2025*). To estimate citrus flower quantities in natural orchards, *Lu et al.* proposed a lightweight recognition model based on an improved YOLOv4 architecture. To reduce model complexity, they employed MobileNetV3 as the backbone feature extractor and incorporated depthwise separable convolutions to accelerate inference (*Lu et al. 2021*).

Li et al. achieved citrus fruit-stem segmentation and precise picking point localization by improving the YOLOv8n-seg model and integrating the CBAM attention mechanism, attaining fruit-stem segmentation accuracy of 97.12%, though generalizability across different varieties remains unverified (*Li et al., 2025*). *Liu et al.* enhanced the YOLOv5s model with the BiFormer attention mechanism to achieve precise detection of citrus spring shoots, identifying 14-4 days before full bloom as the optimal phenological detection window, with a model mAP of 89.6% (*Liu et al., 2025*).

Yin et al. achieved intelligent recognition of citrus shoot growth by improving the YOLOX-Nano algorithm with multi-attention mechanisms, attaining a model mAP of 88.07% and successfully deploying it on mobile terminals, though detection accuracy for fine targets at the germination stage still requires improvement (*Yin et al., 2022*).

In summary, current research on citrus flower, fruit, and shoot recognition has achieved significant progress in technical methodologies. Improved YOLO series models combined with attention mechanisms have effectively enhanced recognition accuracy in complex scenes, research targets have expanded from single organs to multi-phenological dynamic monitoring, and model lightweighting has laid the foundation for real-time orchard applications. However, existing research still faces the following critical challenges. First, current research on citrus target detection predominantly focuses on a single category, such as detecting fruits, flowers, or shoots separately, while comprehensive methods that simultaneously encompass all three categories remain scarce. Second, the adaptability of existing models to complex environments, which are characterized by drastic illumination changes and severe occlusion, still requires further improvement. Third, the lack of standardized datasets hinders horizontal comparison among different studies and impedes the reproducibility verification of research results.

To address the aforementioned limitations, this study performs a targeted expansion of the existing citrus dataset on the PaddlePaddle platform, with particular emphasis on enhancing sample coverage under drastic illumination changes and severe occlusion. Furthermore, a multi-target detection model is constructed to simultaneously encompass three categories: fruits, flowers, and shoots. Specifically, this model introduces the BAM attention mechanism into the neck network of YOLOv10 to enhance feature extraction capability for small-target objects under complex backgrounds, while replacing the original CIoU loss function with the GIoU loss function to improve localization accuracy for densely occluded targets by optimizing geometric alignment between predicted and ground-truth boxes. This improvement scheme aims to enhance the model's simultaneous detection capability for flowers, fruits, and shoots in complex orchard environments through synergistic optimization of feature representation and localization loss, while improving its adaptability to challenging scenarios such as illumination variations and occlusion interference. Furthermore, by constructing a standardized dataset covering multiple phenological stages and varieties, this research provides a foundation for horizontal comparison and reproducibility verification in subsequent studies, thereby promoting the development of citrus organ detection technology toward practical and standardized applications.

MATERIALS AND METHODS

Image dataset

The recognition of citrus flowers, fruits, and shoots is of great significance for cultivation monitoring and yield estimation. The generalization ability and stability of the model heavily rely on citrus datasets covering different growth periods, lighting conditions, and occlusion scenarios. To ensure the effective application of the model under actual agricultural production conditions, the citrus dataset should include images of varying sizes and resolutions, as well as diverse scenarios such as overlap, complex backgrounds, and varying lighting conditions. This study adopts the following combined dataset:

Public dataset: A citrus flower, fruit, and shoot dataset provided by the PaddlePaddle platform, consisting of 420 JPG format images with a resolution of 608×608 pixels. The dataset includes annotation files in YOLO format, covering images from different growth stages that closely resemble real orchard scenes.

Self-constructed dataset: A citrus flower, fruit, and shoot dataset constructed through field photography using a commercial smartphone with a 64MP autofocus camera. This dataset comprises 380 JPG format images with a resolution of 640×640 pixels, encompassing citrus flowers, fruits, and shoots under various lighting conditions, occlusion, overlap, and different levels of background complexity. Sample images are shown in Figure 1.

a. Front lighting b. Back lighting c. Overlap d. Branch occlusion

Fig. 1 - Sample images from the dataset

To simulate the photometric and geometric distortions commonly observed in field operations, data augmentation techniques were applied to the dataset, including horizontal flipping, rotation, cropping, and brightness transformation. Ultimately, the number of sample images was increased to 1,000. The effect of data augmentation is shown in Figure 2.

a. Original image b. Mirror flip c. Rotation d. Cropping e. Brightness

Fig. 2 - Results of data augmentation

To ensure scientific training and testing of the model, the dataset consisting of 1,000 images was divided into training set and test set at a ratio of 8:2. To maintain an equal number of categories in both the training set and test set, stratified sampling was adopted. The number of labels for citrus flowers, fruits, and shoots in the dataset was statistically analyzed, and the label composition is shown in Table 1.

Table 1

Dataset Composition

Type	Data set	Flower	Fruit	Shoot
Train set	800	1813	1929	1351
Test set	200	465	483	338
Sum	1000	2278	2412	1689

Field Deployment and On-site Validation Pipeline

The proposed model was deployed on a handheld mobile terminal integrated with a monocular RGB camera, simulating a real-time in-field detection scenario. The system was calibrated under varying natural illumination conditions between 9:00 AM and 4:00 PM to cover extreme backlighting and shading conditions across different growth stages of citrus. Validation was conducted in a commercial citrus orchard located in Shanxi Juxin Modern Agricultural Demonstration Park. To ensure diversity, samples (fruits, flowers, and shoots) were collected from peripheral (outer canopy) and inner canopy layers, as well as from low-hanging and upper branches, to represent the occlusion complexity in real harvesting environments. Images were captured at a resolution of 1920 × 1080 with a fixed working distance of 0.5 m to 1.5 m. A total of 200 images were collected, consisting of 52 images of flowers, 57 images of fruits, and 99 images of shoots. Due to the difficulty of manual labeling in dynamic outdoor backgrounds, a pre-labeling correction strategy was adopted: initial bounding boxes were generated by the model and then manually corrected by three expert agronomists using LabelImg to establish the high-confidence ground truth for the validation set.

Improved YOLOv10 Model

YOLOv10 is a new generation real-time object detection model proposed by the research team at Tsinghua University in 2024 (*Wang et al., 2024*). Building upon the classic "backbone-neck-head" three-stage architecture inherited from the YOLO series, it achieves breakthroughs in detection performance and computational efficiency through two core innovations.

First, YOLOv10 proposes a consistent dual label assignment strategy, deploying both a one-to-many head (providing rich supervision signals) and a one-to-one head (simulating inference output) during training, while retaining only the latter during inference.

This fundamentally eliminates the reliance on non-maximum suppression (NMS) post-processing, significantly reducing inference latency and simplifying the deployment process. Second, it systematically optimizes the architecture from both efficiency and accuracy perspectives. In terms of efficiency, it adopts lightweight classification heads, spatially-channel decoupled downsampling, and rank-guided module design to reduce computational redundancy. In terms of accuracy, it introduces large-kernel depthwise convolutions in deeper stages to expand the receptive field, and designs a partial self-attention mechanism to enhance global feature representation capability at minimal computational cost. Based on these designs, YOLOv10 achieves industry-leading detection performance while significantly reducing inference latency and computational overhead. The network structure is shown in Figure 3.

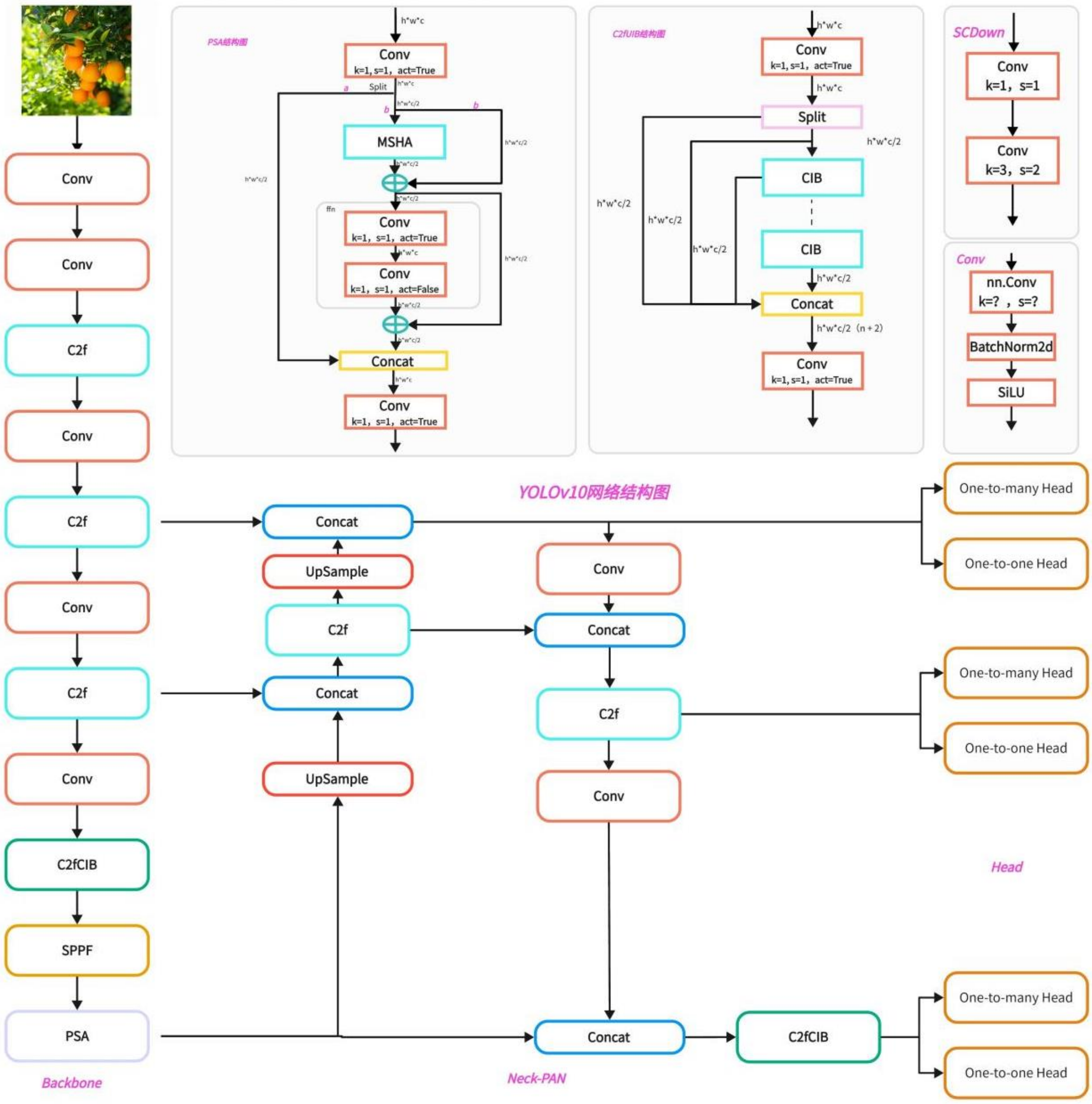


Fig. 3 - Structure of the YOLOv10 network

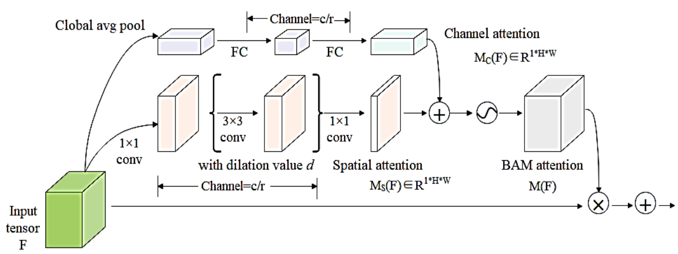


Fig. 4 - BAM Calculation mode

In this paper, the BAM (Bottleneck Attention Module) attention mechanism is incorporated into the neck network of the YOLOv10 model. BAM is an attention mechanism that combines both spatial and channel dimensions(*Park et al., 2020*). For a given feature map, this module processes it through channel attention and spatial attention branches separately to generate a 3D attention map. The specific workflow and structure are illustrated in Figure 4.

The BAM attention module can be integrated with any convolutional neural network, inferring attention maps along both spatial and channel dimensions to generate a 3D attention map that emphasizes important elements. This module combines two attention branches to reduce computational cost and parameter count, while employing residual learning to facilitate gradient flow.

In object detection tasks, IoU is used to measure the overlap between the predicted and ground truth box. Generalized Intersection over Union (GIoU) (*Rezatofighi et al., 2019*) is a loss calculation method for bounding boxes prediction derived from IoU that is scale-invariant. Compared with IoU Loss optimization part is: adding the influence of non-overlapping regions (in the prediction bounding box and the ground truth box there is no overlap part can also reflect the distance between the two boxes), can better measure to the degree of overlap, has a faster convergence speed. The formula is shown below:

$$GIoU = IoU - \frac{A^c - u}{A^c} \tag{1}$$

$$L_{GIoU} = 1 - GIoU \tag{2}$$

As can be seen from the formula, the final result of calculating GIoU Loss is returned as 1-GIoU. Since the value of 1-GIoU is in the range of [0,2] and has a certain “distance” property, i.e., the larger the overlap area between the predicted bounding box and ground truth box, the smaller the loss, and vice versa, the larger it is; and it can avoid the influence of the target shape size and more accurately reflect the interrelationships of the boxes, and possesses scale invariance.

Evaluation metrics

To evaluate the effectiveness of the proposed method, this experiment selects evaluation metrics from multiple dimensions. For model performance evaluation, precision (P), recall (R), mean average precision at IoU threshold 0.5 (mAP50), mean average precision averaged over IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05 (mAP50-95), and the harmonic mean of precision and recall (F1-score) are adopted. For performance efficiency evaluation, floating point operations (FLOPs), number of parameters (Parameters), and weight model memory size (weights size) are included. The specific formulas for each metric are as follows:

$$\mathrm{Presicion} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{5}$$

$$\mathrm{F1\text{-}score} = \frac{2 \bullet \mathrm{Presicion} \bullet \mathrm{Recall}}{\mathrm{Presicion} + \mathrm{Recall}} \tag{6}$$

In the formulas, TP (True Positive) represents the number of correctly predicted targets (both prediction and ground truth are positive), FP (False Positive) represents the number of incorrectly predicted targets (ground truth is negative but prediction is positive), TN (True Negative) represents the number of non-target regions correctly detected as non-targets (both prediction and ground truth are negative), and FN (False Negative) represents the number of missed targets (ground truth is positive but prediction is negative). The F1-score is the harmonic mean of precision and recall.

RESULTS

To comprehensively compare the performance of YOLOv10 models in citrus flower, fruit, and shoot detection, this study selected four models—YOLOv5s (*Jocher et al., 2021*), YOLOv7s (*Wang et al., 2023*), YOLOv8s (*Jocher et al., 2023*), and YOLOv10s—along with the improved YOLOv10s model for comparative experiments. Each model was trained and evaluated on the same dataset with identical training parameters, ensuring the comparability of test results. The baseline detection performance of each model is presented in Table 2.

Table 2

Comparison of Efficiency Performance Among Different Models

model	Params/M	FLOPs/G	VRAM/MB	F1-Score
YOLOv5s	7.0	15.8	159	0.797
YOLOv7s	13.0	33.1	180	0.831
YOLOv8s	11.1	28.4	145	0.793
YOLOv10s	8.0	24.5	127	0.814
YOLOv10s-BAM	8.2	25.2	127	0.827
YOLOv10s-BAM-GIoU	8.2	25.8	127	0.842

To comprehensively evaluate the effectiveness of the improvement strategy, this study conducted a quantitative comparative analysis of the detection performance of different models. In terms of model complexity, the original YOLOv10s has 8.0M parameters, 24.5G FLOPs, and 127MB VRAM usage, while the improved YOLOv10s-BAM-GIoU model has 8.2M parameters, 25.8G FLOPs, and maintains 127MB VRAM usage, indicating that the introduction of the BAM attention mechanism and GIoU loss function brings only minimal computational overhead without imposing additional burden on VRAM usage. In terms of detection accuracy, the original YOLOv10s achieves an F1-score of 0.814, outperforming YOLOv5s (0.797) and YOLOv8s (0.793), but lower than YOLOv7s (0.831). After introducing the BAM attention mechanism, the F1-score of YOLOv10s-BAM increases to 0.827. With the further integration of the GIoU loss function, the F1-score of YOLOv10s-BAM-GIoU reaches 0.842, representing a 2.8 percentage point improvement over the original YOLOv10s and a 1.1 percentage point improvement over YOLOv7s, achieving the optimal performance among all compared models. These results demonstrate that the BAM attention mechanism effectively enhances the model's feature extraction capability for small target organs under complex backgrounds, while the GIoU loss function improves the localization accuracy of densely occluded targets by optimizing the geometric alignment between predicted and ground-truth boxes. The synergistic effect of these two improvements enables the YOLOv10s-BAM-GIoU model to achieve significant detection accuracy gains with minimal computational overhead, validating the effectiveness and superiority of the proposed improvement strategy in complex agricultural scenarios.

To comprehensively evaluate the effectiveness of the proposed improvement strategy, comparative experiments were conducted on six models using a self-constructed citrus flower, fruit, and shoot dataset, with mAP@0.5, mAP@0.5-0.95, and per-class average precision as evaluation metrics. The comparison results of accuracy metrics are shown in Table 3.

Table 3

Comparison of Accuracy Performance Among Different Models

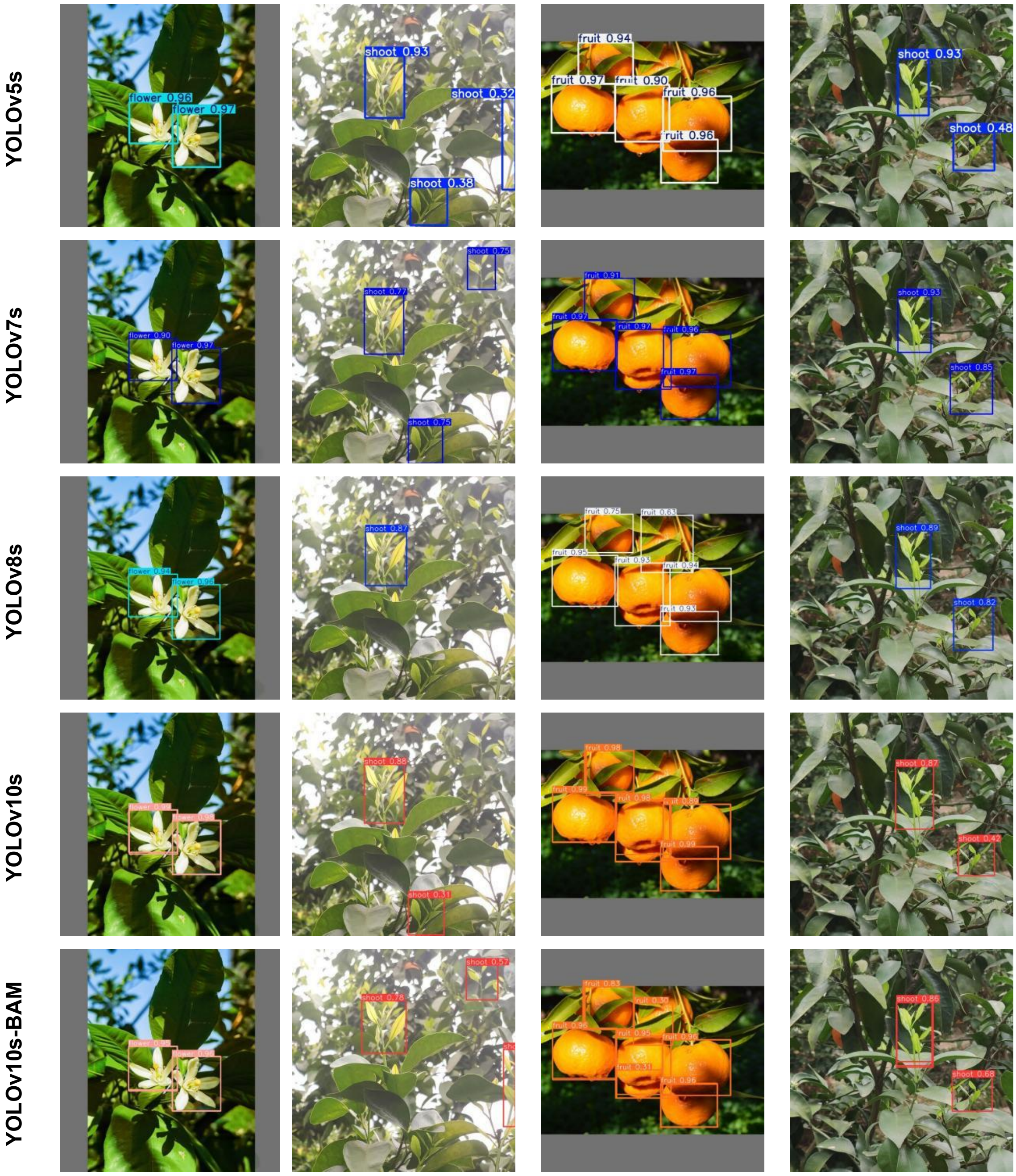
Model	mAP@0.5	mAP@0.5-0.95	Flower	Fruit	Shoot
YOLOv5s	0.831	0.521	0.806	0.782	0.904
YOLOv7s	0.862	0.541	0.864	0.81	0.912
YOLOv8s	0.842	0.539	0.839	0.779	0.909
YOLOv10s	0.796	0.483	0.810	0.724	0.853
YOLOv10s-BAM	0.854	0.539	0.849	0.804	0.908
YOLOv10s-BAM-GIoU	0.891	0.576	0.912	0.836	0.925

The original YOLOv10s achieved mAP@0.5 of 79.6% and mAP@0.5-0.95 of 48.3%, underperforming compared to YOLOv5s (83.1%/52.1%) and YOLOv7s (86.2%/54.1%), likely due to its lightweight design limiting feature capture capability for small targets in complex agricultural scenes. After introducing the BAM attention mechanism, YOLOv10s-BAM achieved mAP@0.5 of 85.4% (a 5.8 percentage point improvement) and mAP@0.5-0.95 of 53.9% (a 5.6 percentage point improvement), demonstrating that the BAM module enhances feature extraction for small target organs under complex backgrounds through parallel channel-spatial attention. With the further integration of the GIoU loss function, YOLOv10s-BAM-GIoU achieved optimal performance among all compared models, with mAP@0.5 reaching 89.1% (a 3.7 percentage point improvement over YOLOv10s-BAM) and mAP@0.5-0.95 reaching 57.6% (a 3.7 percentage point improvement). Compared to the original YOLOv10s, mAP@0.5 improved by 9.5 percentage points and mAP@0.5-0.95 by 9.3 percentage points; compared to the baseline optimal model YOLOv7s, improvements of 2.9 and 3.5 percentage points were achieved, respectively.

In terms of per-class detection accuracy, the original YOLOv10s achieved 81.0%, 72.4%, and 85.3% for flowers, fruits, and shoots, respectively, with relatively lower accuracy for fruits possibly due to dense overlap and similar color to the background. After introducing the BAM attention mechanism, detection accuracy for the three categories improved to 84.9%, 80.4%, and 90.8%, respectively.

With the further integration of the GIoU loss function, YOLOv10s-BAM-GIoU achieved 91.2%, 83.6%, and 92.5% for flowers, fruits, and shoots, representing improvements of 10.2, 11.2, and 7.2 percentage points over the original YOLOv10s, and 4.8, 2.6, and 1.3 percentage points over YOLOv7s, respectively. Notably, the most significant improvements were observed for small targets (flowers and fruits), validating the effectiveness of the proposed improvement strategy for small target recognition in complex scenes. These results demonstrate that while the original YOLOv10s has certain limitations in complex agricultural scenarios, the introduction of the BAM attention mechanism and GIoU loss function effectively compensates for these shortcomings. The synergistic effect of BAM enhancing feature representation for small target organs and GIoU optimizing localization accuracy enables YOLOv10s-BAM-GIoU to achieve significant detection performance improvements with minimal increases in parameters and computational overhead, validating the effectiveness and superiority of the proposed improvement strategy for citrus flower, fruit, and shoot recognition tasks.

To validate the effectiveness of the proposed model in complex environments, the six models in this study were visualized under front-lighting, back-lighting, occlusion, and background interference conditions respectively. The comparison results are show in Figure 5.



YOLOv10s-BAM-GIoU

front-lighting back-lighting occlusion background interference

Fig. 5 - Visualization of model recognition results under different scenes

In the comprehensive detection task of citrus flower, fruit, and shoot recognition, the detection performance of each model under different lighting and occlusion conditions showed significant differences. Under front-lighting conditions, YOLOv5s, YOLOv7s, YOLOv8s, YOLOv10s, YOLOv10s-BAM, and YOLOv10s-BAM-GIoU all demonstrated good detection performance, accurately identifying various citrus organs. Under back-lighting conditions, although YOLOv5s, YOLOv8s, and YOLOv10s could capture citrus shoots at image edges, they exhibited varying degrees of missed detection for small targets; YOLOv7s and YOLOv10s-BAM-GIoU showed stronger small-target recognition capabilities, though missed detections still occurred for targets at image boundaries.

Detection results under occlusion conditions further revealed performance differences among the models. For scenarios with partial leaf occlusion and large-area fruit occlusion, YOLOv5s, YOLOv7s, and YOLOv10s all exhibited varying degrees of missed detection; YOLOv8s performed poorly only under large-area occlusion; YOLOv10s-BAM detected all targets, but annotation completeness required improvement; YOLOv10s-BAM-GIoU performed optimally under occlusion scenarios, achieving the highest detection rate and better annotation accuracy. Under complex background interference conditions, all models demonstrated relatively stable detection performance.

To evaluate model applicability under real agricultural conditions, YOLOv10s and YOLOv10s-BAM-GIoU were validated using 200 unaugmented images from Shanxi Juxin Park, covering backlighting, occlusion, fruit overlap, and tiny shoots without augmentation. The results are presented in Table 4.

Table 4

Validation Results of YOLOv10s and YOLOv10s-BAM-GIoU on The Field Dataset

Model	mAP@0.5	mAP@0.5-0.95	Flower	Fruit	Shoot
YOLOv10s	0.790	0.498	0.741	0.883	0.746
YOLOv10s-BAM-GIoU	0.812	0.519	0.771	0.907	0.757

The proposed YOLOv10s-BAM-GIoU achieves a mAP@0.5 of 81.2% on the challenging field dataset, representing a 2.2% absolute improvement over the standard YOLOv10s (79.0%). Moreover, the proposed model demonstrates its most substantial improvement over YOLOv10s on the "Flower" category (a 3.0 percentage points improvement, from 74.1% to 77.1%), followed by "Fruit" (a 2.4 percentage points improvement) and "Shoot" (a 1.1 percentage points improvement). These enhancements confirm that the integration of the BAM attention mechanism and GIoU loss function effectively mitigates environmental disturbances, particularly in scenarios with complex backgrounds and variable illumination. An unexpected yet insightful finding from the field validation is the superior detection performance for ''Fruit'' (AP = 90.7%) compared to ''Flower'' (77.1%) and ''Shoot'' (75.7%). This contrasts with the common assumption that fruits, being larger and more prone to occlusion, would exhibit the greatest performance degradation. This phenomenon is attributed to two principal factors: (i) the spherical shape and distinctive orange coloration of citrus fruits generate strong contrast against the green foliage background, facilitating robust feature extraction even under variable illumination; and (ii) the field validation images predominantly captured fruits at mid-to-late developmental phases with prominent visual saliency, whereas flowers and shoots presented comparatively greater scale- and color-ambiguity challenges. These findings underscore the importance of category-specific optimization in agricultural vision systems, particularly for enhancing small-target and low-contrast detection.

Comprehensive test results indicate that YOLOv10s-BAM-GIoU achieved the highest detection accuracy and annotation completeness under the most challenging back-lighting and occlusion scenarios. This model not only accurately identifies small targets but also possesses robust capabilities for handling occluded targets, demonstrating its advanced nature in complex agricultural production scenarios.

In conclusion, YOLOv10s-BAM-GIoU can be considered the optimal choice for citrus flower, fruit, and shoot recognition tasks, particularly suitable for complex environmental scenarios commonly encountered in agricultural production.

CONCLUSIONS

This study proposed an improved YOLOv10s model incorporating the BAM attention mechanism and GIoU loss function for the accurate recognition of citrus flowers, fruits, and shoots in complex orchard environments. The experimental results demonstrate that the proposed YOLOv10s-BAM-GIoU model achieves superior detection performance, with an mAP50 of 89.1%, representing improvements of 2.9%~9.5% over the original YOLOv10s and other comparative models. In fine-grained category detection, the model attains mAP50 of 91.2%, 83.6%, and 92.5% for shoots, flowers, and fruits, respectively. The BAM attention mechanism enhances feature extraction capability for small target organs under complex backgrounds through parallel channel-spatial attention branches, while the GIoU loss function improves localization accuracy for densely occluded targets by optimizing geometric alignment between predicted and ground-truth boxes. The synergistic effect of these two improvements enables significant detection accuracy gains with minimal computational overhead, as evidenced by the model's parameter count of 8.2M, computational load of 25.8G FLOPs, and detection speed of 23.6 ms per frame, meeting real-time detection requirements for agricultural applications.

The comparative analysis of six models under varying environmental conditions further validates the robustness and practical applicability of the proposed approach. Under challenging scenarios such as back-lighting and severe occlusion, YOLOv10s-BAM-GIoU consistently outperformed other models, achieving the highest detection rates and annotation accuracy. While YOLOv5s, YOLOv7s, YOLOv8s, and the original YOLOv10s exhibited varying degrees of missed detections for small targets and occluded objects, the improved model demonstrated reliable performance in detecting targets at image boundaries, handling partial and large-area occlusion, and maintaining accuracy under complex background interference. These results confirm that the integration of attention mechanisms and optimized loss functions effectively addresses the key challenges in agricultural target detection, including scale variation, background clutter, and dense occlusion, thereby providing a robust solution for real-world orchard monitoring applications.

The field validation results confirm the model's robustness under realistic orchard conditions, achieving a mAP@0.5 of 81.2% with consistent improvements over the baseline YOLOv10s across all categories, with gains of 3.0 percentage points for "Flower", 2.4 percentage points for "Fruit", and 1.1 percentage points for "Shoot". Notably, an unexpected finding emerged: "Fruit" achieved the highest detection accuracy (AP = 90.7%), substantially surpassing "Flower" (77.1%) and "Shoot" (75.7%). The superior performance of "Fruit" is likely attributable to its higher visual saliency in field images, offering valuable insights for the design of practical agricultural vision systems.

In conclusion, this research contributes to the advancement of precision agriculture by presenting an efficient and accurate deep learning model for citrus organ detection. The YOLOv10s-BAM-GIoU model achieves an optimal balance between detection accuracy and computational efficiency, making it suitable for deployment in field real-time detection systems. The findings highlight the importance of synergistic architectural improvements—combining attention mechanisms for enhanced feature representation with optimized loss functions for precise localization—in addressing the specific challenges of agricultural computer vision tasks. Although the accuracy is somewhat reduced due to the field environment, the system still demonstrates sufficient robustness in specific scenarios, providing a practical reference for the subsequent intelligent management of citrus orchards. Future work could focus on expanding the dataset to include more diverse citrus varieties and growing regions, further optimizing the model for edge device deployment, and extending the approach to other fruit crops and phenological stages. Additionally, integrating the detection system with automated agricultural equipment for precision management practices such as targeted spraying, thinning, and harvesting could realize the full potential of this technology in smart agriculture applications.

ACKNOWLEDGEMENT

This work was supported by Fundamental Research Program of Shanxi Province (Grant No. 202303021212120).

REFERENCES

[1] El Akrouchi, M., Mhada, M., Bayad, M., Hawkesford, M. J., & Gérard, B. (2025). AI-based framework for early detection and segmentation of green citrus fruits in orchards. *Smart Agricultural Technology*, 10, 100834.

[2] Huang, Z., Li, X., Fan, S., Liu, Y., Zou, H., He, X., Xu, S., Zhao, J., & Li, W. (2025). ORD-YOLO: A Ripeness Recognition Method for Citrus Fruits in Complex Environments. *Agriculture*, 15(15), 1711.

[3] Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics (Version 8.0.0) [*Computer software*]. https://github.com/ultralytics/ultralytics.

[4] Jocher, G., Stoken, A., Borovec, J., et al. (2021). ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support. *Zenodo*.

[5] Li, H., Yin, Z., Zuo, Z., Pan, L., & Zhang, J. (2025). Precision citrus segmentation and stem picking point localization using improved YOLOv8n-seg algorithm. *Frontiers in Plant Science*, 16, 1655093.

[6] Lin, Y., Hu, W., Zheng, Z., & Xiong, J. (2023). Citrus identification and counting algorithm based on improved YOLOv5s and DeepSort. *Agronomy*, 13(7), 1674.

[7] Liu, W. H., Jian, C., Zhang, S. A., Cong, G. X., Wang, H. B., Li, J. Y., & Zheng, Y. Q. (2025). Enhanced YOLOv5s with BiFormer attention for citrus spring shoot detection: Optimized phenological period and cross-regional application. *Smart Agricultural Technology*, 101446.

[8] Lu, J., Lin, W., Chen, P., Lan, Y., Deng, X., Niu, H., ... Luo, S., (2021). Research on lightweight citrus flowering rate statistical model combined with anchor frame clustering optimization. *Sensors*, 21(23), 7929.

[9] Lyu, Q., Lin, G., Jiang, J., Wang, M., Zhang, H., & Yi, S. (2024). Detecting green citrus fruit in natural scenes using improved YOLOv5s model. *Transactions of the Chinese Society of Agricultural Engineering*, 40(18), 147-154.

[10] Park, J., Woo, S., Lee, J. Y., & Kweon, I. S. (2020). A simple and light-weight attention module for convolutional neural networks. *International Journal of Computer Vision, 128*(4), 783-798.

[11] Qin, X., Huang, R., & Hua, B. (2021, March). Research and implementation of yield recognition of Citrus reticulata based on target detection. In *Journal of physics: conference series.* Vol. 1820, No.1, p.012128). IOP Publishing.

[12] Qiu, X., Zhu, F., Huang, P., Chen, H., Li, Y., Pu, C., & Wang, S. (2025). Quantifying high-temperature-induced reproductive growth imbalance in citrus at anthesis: Insights from the CF-ASPM model. *Computers and Electronics in Agriculture*, 236, 110421.

[13] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 658-666.

[14] Subeesh, A., Kumar, S. P., Chakraborty, S. K., Upendar, K., Chandel, N. S., Jat, D., ... Khan, M. M., (2024). UAV imagery coupled deep learning approach for the development of an adaptive in-house web-based application for yield estimation in citrus orchard. *Measurement*, 234, 114786.

[15] Ünal, İ., Eceoğlu, O., (2025). A Lightweight Instance Segmentation Model for Simultaneous Detection of Citrus Fruit Ripeness and Red Scale (Aonidiella aurantii) Pest Damage. *Applied Sciences*, 15(17), 9742.

[16] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *Advances in neural information processing systems*, 37, 107984-108011.

[17] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464-7475).

[18] Xiao, X., Wang, Y., Jiang, Y., Wu, H., Zhang, Z., & Wang, R. (2024). AC-YOLO: citrus detection in the natural environment of orchards. *Journal of Agricultural Engineering*, 55(4).

[19] Xu, L., Zhu, S., Chen, X., Wang, Y., Kang, Z., Huang, P., & Peng, Y. (2020). Citrus recognition in real scenarios based on machine vision. *DYNA-Ingeniería e Industria*, 95(1).

[20] Yang, C. H., Xiong, L. Y., Wang, Z., Wang, Y., Shi, G., Kuremot, T., ... & Yang, Y. (2020). Integrated detection of citrus fruits and branches using a convolutional neural network. *Computers and Electronics in Agriculture*, 174, 105469.

[21] Yin, X. B., Deng, X. L., Lan, Y. B., & Chen, X. (2022). Intelligent recognition of citrus shoot growth based on improved YOLOX-Nano algorithm. *Journal of South China Agricultural University*, 44(1), 142-150.