Evaluation of lychee winter shoot length using UAV remote sensing technology

Zifan Shen[1,2,3,4], Baihan Liu[1], Rui Xu[1,2,3,4], Yiwei Wang[1,2,3,4], Heguang Sun[1,2,3,4], Yongshun Liu[5], Yubin Lan[1,2,3,4], Xiangbao Meng[6], Xiaoling Deng[1,2,3,4*]

(1. *College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China*;
2. *National Center for International Collaboration Research on Precision Agricultural Aviation Pesticide Spraying Technology, Guangzhou 510642, China*;
3. *Center for International Cooperation and Disciplinary Innovation of Precision Agricultural Aviation Applied Technology ('111 Center'), Guangzhou 510642, China*;
4. *Guangdong Provincial Engineering Research Center for Smart Agriculture, Guangzhou 510642, China*;
5. *College of Mathematical, Physical and Computational Sciences (College of Climate Change and Artificial Intelligence), University of Reading, RG6 6UR, UK*;
6. *Guangzhou Joinken Network Technology Development Co., Ltd., Guangzhou 510630, China*)

Abstract: Lychee is an important cash crop in southern China. The excessive growth of winter shoots in the early winter season will lead to an increase in nutrient consumption, which in turn affects flower bud differentiation and fruit yield. To address the issue of low efficiency in traditional manual measurement methods, this study proposes an automated detection method using UAV remote sensing technology and an improved YOLOv8n_OBB_SEB algorithm. Through multi-dimensional optimization, this method successfully solves the issue of the small size of winter shoots, similar color to branches, and leaf occlusion in the orchard environment. The specific improvements include: using the SAHI algorithm for image slicing to assist inference to improve the recognition ability of small targets; embedding the Starblock in the StarNet model into the C2f module and replacing the original C2f module in the Backbone, which reduces the number of parameters and strengthens the feature extraction ability; replacing the Concat module in the Neck part with the BiFPN structure to optimize multi-scale feature fusion; introducing the EMA attention mechanism and embedding it into the C2f module in the Neck part to achieve pixel-level attention allocation and enhance the distinguishability between the target and the background. The experimental results show that on the lychee winter shoot test set, the detection accuracy of the improved YOLOv8_OBB_SEB algorithm reaches 89.2%, which is 20.7% higher than that of the original YOLOv8_OBB algorithm. Compared with other mainstream algorithms, YOLOv8_OBB_SEB shows stronger competitiveness and robustness. Through inference detection, the four coordinates of the target rotation box can be obtained, and the actual size can be calculated by converting the pixel height to estimate the real length of the lychee winter shoots. According to the estimation results, this paper divides the winter shoots into two groups: those requiring drug intervention and those not requiring drug intervention. The specific judgment standard is that when the length of the winter shoot exceeds 3 centimeters, it is classified into the group requiring drug intervention, and when the length of the winter shoot is less than 3 centimeters, it is classified into the group not requiring drug intervention. Remote sensing data of 24 lychee trees were collected on December 3, 2024. The spraying requirements were determined through manual field surveys, which were then compared and verified with the model inference results. Finally, it was concluded that the accuracy of the model reached 83.3%. This classification method provides reliable decision support and a clear decision-making basis for the precise management of winter shoots.

Keywords: small object detection, UAV remote sensing, computer vision, deep learning
DOI: 10.25165/j.ijabe.20251806.9852

Citation: Shen Z F, Liu B H, Xu R, Wang Y W, Sun H G, Liu Y S, et al. Evaluation of lychee winter shoot length using UAV remote sensing technology. Int J Agric & Biol Eng, 2025; 18(6): 241–249.

1 Introduction

Lychee is a subtropical fruit native to southern China and is also an important agricultural product in the region. Around the turn of autumn and winter, if temperatures are abnormally high, lychee trees tend to sprout winter shoots. The growth of these winter shoots not only consumes a large amount of nutrients but also interferes with flower bud differentiation, ultimately leading to reduced lychee yields. Based on research on relevant literature and guidance from local planting experts, the determined response plan is: when over

Received date: 2025-04-18 **Accepted date:** 2025-09-03
Biographies: Zifan Shen, MS, research interest: UAV remote sensing, Email: 3254379677@qq.com; **Baihan Liu**, Undergraduate Student, research interest: UAV remote sensing, Email: 2580858472@qq.com; **Rui Xu**, MS, research interest: UAV remote sensing, Email: 1346615818@qq.com; **Yiwei Wang**, MS, research interest: UAV remote sensing, Email: 33748968@qq.com; **Heguang Sun**, PhD candidate, research interest: UAV Remote Sensing Inversion, Email: 17746317626@163.com; **Yongshun Liu**, MS, research interest: UAV remote sensing, Email: liuyongshun2024@163.com; **Yubin Lan**, PhD, research interest: Precision Agricultural Aviation, Email: ylan@scau.edu.cn; **Xiangbao Meng**, PhD, research interest: Agricultural Informatization and Intelligent Equipment Technology: Research and Application, Email: mengxb@e-jiankun.com.
***Corresponding author: Xiaoling Deng**, PhD, Professor, research interest: UAV remote sensing and Computer Vision, College of Electronic Engineering South China Agricultural University, Guangzhou 510620, China, Tel: 13570521308, Email: dengxl@scau.edu.cn.

15% of the winter shoots on lychee trees exceed 3 cm in length, it is necessary to spray a 1250–2500 times diluted 40% ethephon aqueous solution, combined with a 300–500 times diluted paclobutrazol wettable powder[1-3]. Therefore, monitoring the length of winter shoots is crucial for effective orchard management and yield enhancement. However, detecting new lychee shoots presents several challenges. First, the winter shoots are very small, making it difficult to observe and measure them directly. Secondly, the dense branches and leaves of the lychee tree often obscure the winter shoots, further complicating the detection process. This makes traditional manual detection methods not only inefficient but also prone to obstruction and perspective issues, resulting in inaccurate measurements. Therefore, effective detection of lychee winter shoots remains a technical problem that needs to be addressed in orchard management.

Using UAVs to capture high-resolution images of crops and employing deep learning[4] and machine learning techniques for crop growth recognition and diagnosis is currently a mainstream research method. Many scholars have applied this approach to problems such as citrus flower counting[5], disease recognition[6], and canopy segmentation[7]. Huang et al.[8] integrated the CBAM attention module and ASFF module with YOLOv5s, improving the model's accuracy in detecting whiteflies. Li et al.[9] combined drone remote sensing with image segmentation, integrating ResNet34, CBAM, and U-Net to achieve precise segmentation of the lychee tree canopy, with a segmentation accuracy of 90.98%. Song et al.[10] combined the RESAM module with YOLOv8n to propose the SEYOLOX-tiny model, successfully extracting maize ears from drone-captured maize field images, achieving excellent detection accuracy with a mAP0.5 index of 95.0%. Qing et al.[11] developed YOLOPC to detect citrus diseases, with approximately 75% fewer parameters than YOLOv5 Nano, achieving an accuracy of 94.5%, and linked GPT for auxiliary diagnosis. Liang et al.[12] integrated Swin-transformer, BiFPN, CBAM, and CIOU into the YOLOv5 model for detecting lychee winter shoots, achieving an accuracy of 79.56%. However, this approach did not address the estimation of the length of the lychee winter shoots and the determination of whether pesticide spraying is required.

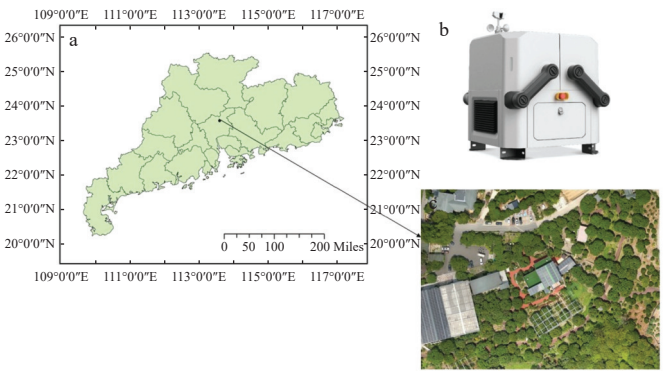
In recent years, deep learning technologies have made significant advancements in object detection, yet small object detection and occlusion remain major challenges. There are two primary reasons for the poor performance of small object detection: First, small objects are significantly fewer in number compared to medium and large objects in datasets, resulting in insufficient feature learning for small objects. Second, small objects occupy fewer pixels, and after multiple convolution and pooling operations, their pixel information becomes excessively sparse, which can even lead to the loss of small objects. To address these issues, optimizations can be made in two aspects: the data preprocessing stage and the model architecture. Kisantal et al.[13] proposed a method for oversampling small object images by repeatedly copying, pasting, and enhancing these images. To tackle the problem of limited pixels in small objects, Noh et al.[14] proposed a feature-level super-resolution approach with supervisory signals using generative adversarial networks. Akyon et al.[15] developed the SAHI algorithm, which improves small object detection accuracy through image slicing and auxiliary inference. Zhu et al.[16] also combined transformers and YOLOv5 to detect small objects in drone-captured remote sensing images, achieving strong performance in the VisDrone 2021 Challenge[17]. Cai et al.[18] proposed the PKINet model, which employs non-dilated multi-scale convolution kernels to extract features from targets of various scales and capture local context. Additionally, a Context Anchor Attention module was introduced in parallel to capture long-range contextual information, effectively addressing the issues where large convolution kernels may introduce significant background noise and dilated convolutions may result in overly sparse feature representations. Lim et al.[19] proposed a context-aware object detection method to improve the accuracy of small object detection. Liu et al.[20] combined the multi-scale attention module EMA, path aggregation feature pyramid network PAFPN, MPDIoU, and YOLOv8 to enhance the detection accuracy of small target green plums in complex orchard environments, achieving an accuracy of 92.3%.

In agricultural production, the management of winter shoots in lychee orchards typically involves spraying pesticides when the winter shoots reach 3 cm in length to inhibit their growth. In remote sensing images, these winter shoots are considered small targets. This study aims to develop a detection method more suitable for measuring the length of lychee winter shoots by combining an improved YOLOv8 algorithm with drone remote sensing technology. 1) A remote sensing dataset of lychee winter shoots was constructed through drone data collection. 2) To effectively detect these targets, this paper proposes an improved YOLOv8 algorithm, named YOLOv8n_OBB_SEB. Experimental results show that the accuracy of YOLOv8n_OBB_SEB in detecting lychee winter shoots reached 89.2%, a 20.7% improvement over the original YOLOv8n_OBB algorithm's 68.5%. Furthermore, the proposed YOLOv8_OBB_SEB algorithm outperforms current mainstream detection algorithms on the test set. 3) This study establishes a scientific winter shoot classification standard by inferring shoot length through pixel-based measurement of rotated bounding boxes: shoots exceeding 3 cm require pesticide application, while those under 3 cm need no treatment, providing quantitative decision-making support for precision lychee orchard management.

2 Materials and methods

2.1 Data collection

The dataset used in this paper was sourced from the lychee orchard in Conghua District, Guangzhou, Guangdong Province, China (23.56°N, 113.61°E). The data were collected using a DJI M30T Enterprise Edition drone (Figure 1). The drone is equipped with a camera with a resolution of 12 million pixels (4000×3000), supporting 16× optical zoom and 200× digital zoom. The data collection covers 34 lychee trees of the Jinggang Hongnuo variety, with the time range from late November 2023 to early January 2024



Note: a. Main study area location; b. Acquisition equipment: DJI Airport; c. Sampling point at Li Bo Garden, Conghua, Guangzhou, China.

Figure 1 Collection location and equipment

and December 2024, spanning two years. Images were captured daily from 2:00 p.m. to 4:00 p.m. The flight altitude of the drone was set to 6 m to ensure clear capture of lychee winter shoots while avoiding interference from propeller airflow on the lychee trees. For each lychee tree, three sampling points were set in each of the four cardinal directions (east, south, west, and north), resulting in a total of 12 sampling points per tree. A total of 1100 remote sensing images were collected. Due to varying weather conditions during data collection, the images show differences in light intensity, which enhances the diversity of the dataset and provides strong support for the generalization ability of the model.

2.2 Data pre-processing

Due to factors such as lighting, focus, and camera shake, some of the remote sensing images had low clarity, which led to the removal of images that did not meet the training requirements. We used the roLabelImg annotation tool (as shown in Figure 2), which employs a rotated bounding box framework to more accurately estimate the length of winter shoots. Then, we used a script to convert the XML files into TXT files that met the model's training standards. To improve the detection of small objects, we performed image slicing. In the experiment, each image was divided into 2048×2048 pixel sections with a 20% overlap between adjacent slices. This image slicing not only significantly improved the accuracy of small object detection but also accelerated the GPU's image processing speed. To enhance the model's learning performance and generalization ability in complex orchard environments, data augmentation techniques were applied, such as mixup, mosaic, flipping, and brightness adjustments (as shown in Figure 3), with the effects demonstrated in the figure. These measures effectively improved the model's generalization ability and detection accuracy. In the end, 1100 original images were collected, which were processed into 6600 sub-images through slicing. These images were divided into training, validation, and test sets in an 8:1:1 ratio, with the training set containing 5280 images, and both the validation and test sets containing 660 images each.

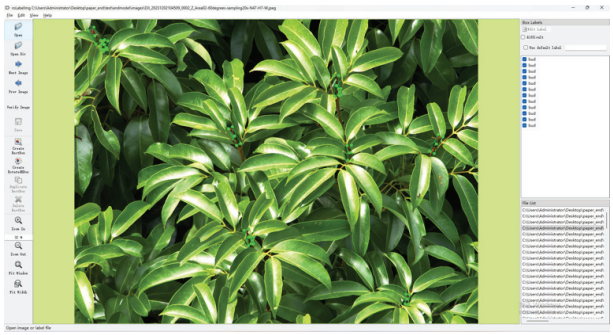


Figure 2 A shoot annotation tool used in the experiment named roLabelImg

a. Horizontal flipping b. Vertical flipping c. Brightness variation d. Mixup

Figure 3 Examples of image enhancement methods

2.3 Proposed model

When lychee winter shoots exceed 3 cm in length, pesticide application is required to inhibit their growth and prevent negative impacts on fruit production in the following year. The dataset of this study incorporated both newly emerged and mature winter shoots, covering various growth stages to ensure model generalizability. As shown in Figure 4, from top to bottom, the last image in each row displays the distribution of the horizontal coordinate X of the rotated bounding box center, the distribution of the vertical coordinate Y of the rotated bounding box center, the distribution of the width of the rotated bounding box, and the distribution of the height of the rotated bounding box. It can be observed that the rotated bounding boxes are evenly distributed across the image, with most bounding boxes having a width less than 0.05 of the total image width and a height less than 0.2 of the total image height. The detection of lychee winter shoots presents three main challenges: 1) Limited features: The low resolution of winter shoots in remote sensing images restricts the extraction of distinctive characteristics; 2) Scale variation: The dataset contains winter shoots of different lengths, exhibiting significant scale differences across growth stages; 3) Background interference: The similar coloration between winter shoots and tree trunks/dry branches makes target-background differentiation difficult for models.

To enhance the performance of YOLOv8n_OBB on the lychee winter shoot dataset, the YOLOv8n_OBB_SEB algorithm was developed. This model improves feature extraction quality by embedding the Starblock module from the StarNet[21] backbone network into the C2f module, replacing the original C2f module. Additionally, the BIFPN module[22] is introduced to replace the Concat module, further enhancing feature fusion. The EMA module[23] is embedded into the C2f module, replacing the C2f module in the neck section, significantly improving the model's ability to identify key feature areas in complex backgrounds. Finally, Non-Maximum Suppression (NMS) is applied to

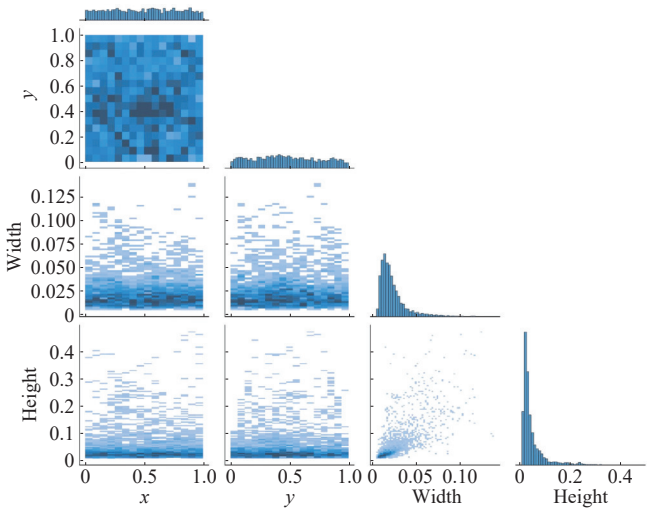


Figure 4 Detailed distribution of annotation boxes

successfully merge the detected sliced sub-images. The model structure is shown in Figure 5. After inference with the YOLOv8n_OBB_SEB model, the actual length of winter shoots is calculated based on rotated bounding box coordinates: If the length is greater than 3 cm, pesticide spraying is required; if the length is less than or equal to 3 cm, no treatment is needed.

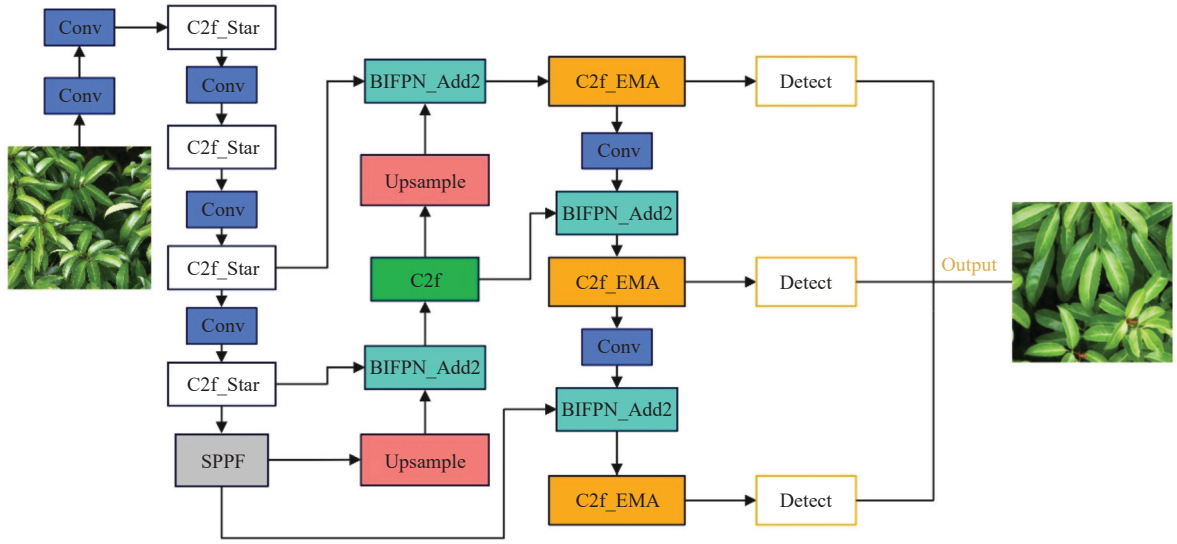


Figure 5 The final proposed YOLOv8n_OBB_SEB network structure

2.3.1 Slicing-aided hyper-inference

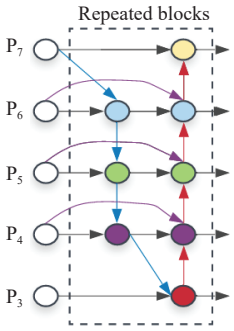
Considering the issues of limited features and small targets, this paper adopts the SAHI algorithm. SAHI is a slicing-assisted hyper-inference algorithm designed for small object detection in ultra-large images. It allows direct application to existing networks without the need for model redesign or retraining, enhancing the accuracy of small object detection. SAHI performs slicing inference on the original image by dividing it into smaller regions for prediction. These prediction results are then merged using Non-Maximum Suppression (NMS). This approach allows for more accurate detection of small objects by focusing on smaller regions of the image. This technique is particularly useful in situations where the target size is small, resulting in limited pixel representation.

2.3.2 Bidirectional Feature Pyramid Network

Considering the issue of scale variation of targets, this study has integrated the BiFPN structure into the YOLOv8n_OBB algorithm. The conventional FPN[24] primarily fuses multi-scale feature maps through a top-down pathway, whereas BiFPN not only retains this top-down pathway but also introduces a complementary bottom-up pathway, establishing bidirectional information flow to generate more enriched feature representations. Furthermore, BiFPN employs learnable weights to perform weighted fusion of input feature maps, as opposed to simple concatenation or summation, enabling the model to automatically learn the relative importance of each feature map and thereby enhance the quality of fused features. The architecture of BiFPN is illustrated in Figure 6. Previous studies[25,26] have demonstrated that integrating the BiFPN module into object detection models yields significant performance improvements.

2.3.3 Efficient Multi-scale Attention

Considering the complex background environment and the high similarity in color and morphology between winter shoots and branches, the detection accuracy in winter shoot identification tasks is often suboptimal. To address this challenge, this study innovatively integrates the EMA module into the C2f module of the YOLOv8n model. The EMA (Efficient Multi-scale Attention) attention mechanism enhances the model's accuracy by establishing both short-term and long-term dependencies using multi-scale parallel subnetworks, without the need for channel reduction. It reshapes part of the channels into the batch dimension and groups the channel dimension into multiple sub-features, ensuring that spatial semantic features are evenly distributed within each feature group. This structure not only retains the information in each channel but also reduces the computational overhead. Additionally, the EMA module recalibrates the channel weights in each parallel branch by encoding global information. It further aggregates the output features of the two parallel branches through cross-dimensional interaction to capture pixel-level pairwise relationships. The EMA network structure is shown in Figure 7.



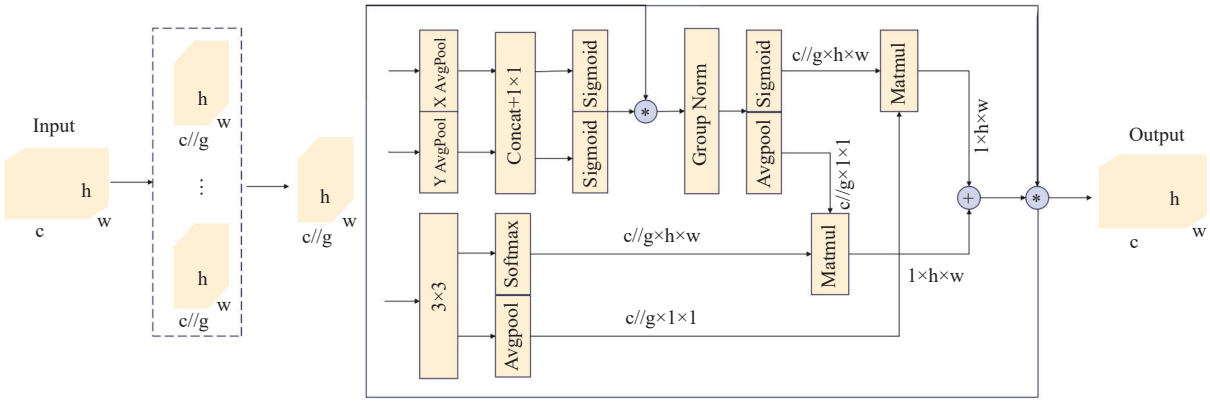
Note: P_3-P_7 are the input features of layers 3-7.

Figure 6 BiFPN structural

2.3.4 StarBlock

Considering the limited features of lychee winter shoots and the adoption of the lightweight Nano version, this study innovatively integrates the StarBlock module (Figure 8) into the C2f module to enhance feature extraction quality.

The Star operation represents a breakthrough from conventional neural networks' channel-number-increment approach. It achieves feature fusion through element-wise multiplication across different feature subspaces, operating similarly to kernel function mechanisms. When stacked in multiple layers, this operation enables exponential expansion of implicit dimensionality, constructing high-dimensional feature spaces with minimal layers.



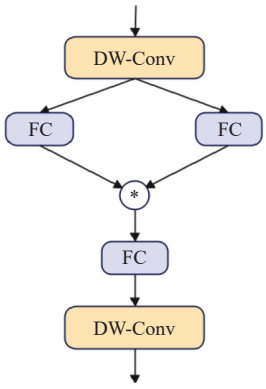
Note: "g" means the divided groups; "c" means the number of channels; "h" means the feature height; "w" means the feature width; "X Avg Pool" represents the 1D horizontal global pooling, and "Y Avg Pool" indicates the 1D vertical global pooling.

Figure 7 EMA structure

This innovation allows YOLOv8n_OBB (a low-dimensional model) to maintain low computational complexity while significantly enhancing feature extraction performance. The star operation can be written as:

$$w_1^T x * w_2^T x = \left(\sum_{i=1}^{d+1} w_1^i x^i\right) * \left(\sum_{j=1}^{d+1} w_2^j x^j\right) = \sum_{i=1}^{d+1}\sum_{j=1}^{d+1} w_1^i w_2^j \pi^i x^j = \alpha_{(1,1)} x^1 x^1 + \ldots + \alpha_{(4,5)} x^4 x^5 + \ldots + \alpha_{(d+1,d+1)} x^{d+1} x^{d+1} \quad (1)$$

where, d is the input channel number, w_1 and w_2 are the weight vectors, x is the feature vector, i and j are the indices of the channels, and α is the coefficient for each term.



Note: *: element-wise mul. (star); DW-Conv: The kernel size of Depthwise Convolution is 7×7, and the stride is 1; FC: Fully Connected Layer.

Figure 8 Starblock structure

2.3.5 Automated Classification of Lychee Winter Shoots

The method for measuring the length of lychee winter shoots in this experiment is as follows: The YOLOv8n_OBB_SEB model is used for object detection to obtain the rotated bounding box of the target, and then the length is calculated based on the coordinates of the four points of the bounding box. Orchard drone inspections are usually conducted by shooting along preset flight paths, so the shooting scheme remains consistent. This experiment uses the DJI M30T drone, which features an image control-free design and is equipped with a lens distortion correction function and a laser rangefinder.

To explore the differences, comparative verification was simultaneously conducted through the automatic adjustment of the DJI Dock and a complete geometric calibration process: Five high-reflectivity PVC ground control panels (GCPs, 30 cm×30 cm, with a black and white checkerboard pattern) were placed in the experimental orchard, and their actual coordinates were recorded using RTK-GPS as the ground truth reference. The DJI M30T drone (RTK nominal accuracy: horizontal 1 cm + 1×10^{-6}, vertical 1.5 cm + 1×10^{-6}) was used for positioning accuracy verification. In the static verification, the drone was stationary above the GCPs, and the calculated horizontal RMSE=±1.2 cm and vertical RMSE=±1.6 cm, which meet the accuracy requirements.

Lens distortion correction was performed using Zhang Zhengyou's[27] method: 20 images of a 12×9 checkerboard (with a square side length of 3 cm) taken from different angles were calibrated using OpenCV to obtain the intrinsic matrix and distortion coefficients, which were then imported into Metashape to complete the correction. The results show that the two schemes are consistent, with a pixel resolution of 0.01 cm per pixel. The actual length of the lychee winter shoots satisfies the following equation:

$$L_{new} = L_{obb} \times U_{px} \quad (2)$$

where, L_{new} represents the shoot length, cm; L_{0bb} represents the rotated bounding box length; U_{px} represents the pixel unit length, cm per pixel.

3 Experimental results and analysis

3.1 Experimental environment

The training environment for this paper consists of Windows 10 with 32 GB of RAM and an NVIDIA GeForce RTX 3090 GPU. Model training was implemented using Python 3.9.0, PyTorch 1.10.1, and CUDA 11.3.

In this study, a model input size of 2048×2048 was selected, with Adam as the optimizer. The batch size was set to 16, and the number of epochs was set to 300. The initial learning rate was 0.001, and an early stopping strategy was applied to prevent overfitting. Commonly used evaluation metrics in object detection, including mean average precision (MAP), model size (Weights), and frames per second (FPS), were employed in this study.

3.2 Evaluation metrics

In object detection tasks, precision (P), recall (R), average precision (AP), model size, and frames per second (FPS) are commonly used as metrics to evaluate the overall performance of the model. The calculations for P, R, and AP are as follows:

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$AP = \frac{p}{num(Total\ Objects)} \quad (5)$$

Where TP represents the number of correctly detected lychee sprouts, FP represents the number of falsely detected lychee sprouts, and FN represents the number of missed lychee sprouts.

3.3 Ablation experiment

To tackle the challenge of detecting the extremely small targets of lychee winter shoots and improve the detection performance, researchers conducted experiments on the YOLOv8n_OBB model using three datasets: the original dataset, the non-overlapping sliced dataset, and the overlapping sliced dataset. When the image data is directly input without slicing, the MAP50 of the model can only reach 68.5%. In contrast, the MAP50s of the non-overlapping sliced dataset and the overlapping sliced dataset are 79.1% and 80.6%, respectively. Moreover, to further enhance the model's performance, this paper integrates three key improved modules, namely StarBlock, BiFPN, and EMA, and seamlessly incorporates them into the YOLOv8n_OBB framework to comprehensively improve the model's performance on the lychee winter shoot dataset. Ablation experiments were conducted for each module, and the specific results are listed in Table 1. The experimental results indicate that embedding StarBlock into the C2f module of the backbone network effectively enhances the model's feature representation ability without increasing model parameters or computational load, leading to a 2.9% improvement in MAP50. Additionally, when the mAP50 value reaches its optimal value, YOLOv8n_OBB requires 248 epochs, while the YOLOv8n_OBB with StarBlock only requires 228 epochs (as shown in Figure 9). Furthermore, the BiFPN module dynamically adjusts feature importance through a weighting mechanism and supports bidirectional feature flow (top-down and bottom-up), enabling efficient multi-scale feature fusion, which improves MAP50 by 1.2%. Meanwhile, embedding the EMA module into the C2f module of the neck network significantly enhances the model's ability to distinguish between target and background regions, effectively improving overall detection performance, with a 2.3% increase in MAP50. By combining the advantages of these three modules, the YOLOv8n_OBB_SEB model achieves a MAP50 of 85.4%, a 4.8% improvement over the original YOLOv8n_OBB. In the lychee winter shoot detection task, the model demonstrates stronger detection capabilities, good robustness, and great potential for practical applications.

Table 1 Ablation experiment

Methods	MAP50/%	*P*/%	*R*/%	Weight/Mb	GFLOPs	FPS
YOLOv8n_OBB	80.6	85.7	76.7	6.74	9.3	158.1
YOLOv8n_OBB+BIFPN	81.8	83.5	80.2	9.33	14.3	138.6
YOLOv8n_OBB+C2f_EMA	82.9	85.5	79.2	9.28	14.4	139.0
YOLOv8n_OBB+C2f_Star	83.5	83.1	82.8	8.39	12.1	123.5
YOLOv8n_OBB+BIFPN+C2f_EMA	83.5	82.4	81.1	9.34	14.5	132.7
YOLOv8n_OBB+C2f_EMA+C2f_Star	80.7	81.4	80.3	8.42	12.3	118.5
YOLOv8n_OBB_SEB(proposed)	85.4	85.2	83.2	8.51	12.7	118.2

3.4 Comparative experiments on attentional mechanisms

To validate the effectiveness of the EMA module, this study conducted comparative experiments with CBAM[28] and LSKA[29] modules (Table 2). Experimental results demonstrate that the EMA module improved mAP50 from 80.6% to 82.9%, achieving a 2.3 percentage point accuracy increase and outperforming other attention mechanisms. The EMA module's multi-scale feature extraction and cross-space interaction mechanisms enable more comprehensive feature capture. Unlike CBAM's separate processing of channel/spatial attention, EMA integrates multi-scale features with cross-space interactions, significantly enhancing detection performance. For small object detection tasks like lychee winter shoot identification, EMA demonstrates superior local feature extraction capabilities compared to LSKA's large kernel decomposition approach.

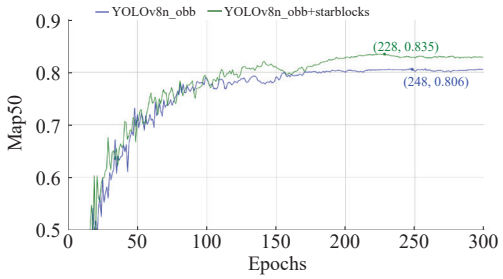


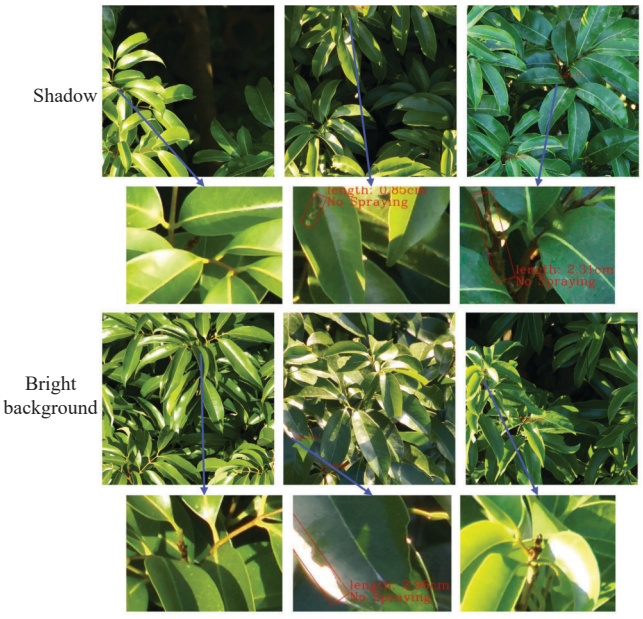
Figure 9 Comparison of MAP50 between YOLOv8n_OBB and YOLOv8n_OBB+StarBlock

Table 2 Comparison of attention mechanisms

Model	Attention Methods	MAP50/%	P/%	R/%
YOLOv8n_obb	-	80.6	85.7	76.7
YOLOv8n_obb	CBAM	81.9	81.9	78.6
YOLOv8n_obb	LSKA	82.8	85.1	79.5
YOLOv8n_obb	EMA	82.9	85.5	79.2

3.5 Performance comparison of different object detection models

Research indicates that small-sample object detection still faces accuracy degradation issues due to background interference[30,31]. To address this, this study optimized the YOLOv8n_OBB model. Experiments revealed that the original model exhibited missed and false detections under strong lighting or shadowed backgrounds (Figure 10). By incorporating image overlapping slicing technology and improving the model architecture (with particular focus on optimizing the attention mechanism and multi-scale feature fusion modules), we developed an enhanced YOLOv8n_OBB_SEB model. As listed in Table 3, the improved model demonstrates outstanding performance in lychee winter shoot detection: achieving 85.4%



Note: Magnified view of the local region.

Figure 10 YOLOv8n_OBB inference results

MAP50, 85.2% precision (*P*), and 83.2% recall (*R*), significantly outperforming mainstream algorithms including R-CNN[32], R3Det[33], S2A-Net[34], ReDet[35], and YOLOv5_OBB. Moreover, the model maintains excellent computational efficiency, with a compact size of 8.51 MB, computational load of 12.8 GFLOPs, and real-time processing speed reaching 118.2 FPS.

Table 3 Comparison of results from different algorithms

Methods	MAP50/%	P/%	R/%	Weight/Mb	GFLOPs	FPS
YOLOv8n_OBB	80.6	85.7	76.7	6.74	9.60	158.1
R-CNN	76.8	76.5	82.5	314	198.5	18.6
R3Det	68.4	76.4	72.4	318	328.7	16.4
S2A-Net	70.9	85.6	77.6	295	196.2	17.2
ReDet	58.1	65.2	66.9	244	54.3	8.2
YOLOv5n_OBB	71.9	71.3	67.4	4.05	4.2	180.8
YOLOv8n_OBB_SEB	85.4	85.2	83.2	8.51	12.8	118.2

3.6 Visualization of detection results

To verify the superiority of the proposed model in lychee winter shoot detection, we randomly selected several images from the test set for a comparative analysis of detection performance. The YOLOv8n_OBB model performed poorly on the lychee winter shoot dataset, mainly due to false positives and missed detections under conditions of bright light and shadow, as shown in Figure 10. Specifically, the model mistakenly identified spots on leaves, mature branches, and light reflections as winter shoots. In contrast, the improved YOLOv8n_OBB_SEB model demonstrated superior performance in distinguishing between the target and background, and it was able to accurately detect winter shoots even in dimly lit environments, as shown in Figure 11. This indicates that the YOLOv8n_OBB_SEB model has significantly improved robustness and accuracy under complex lighting conditions.

3.7 Model generalization evaluation

In the experimental study conducted on December 23, 2024, aerial images of six lychee trees were captured using drone photography, followed by intelligent recognition analysis employing the YOLOv8n_OBB_SEB deep learning model. The winter shoots of the six sampled lychee trees exhibited a gradient distribution in length (classified into short, medium, and long types). In practical agricultural production, pesticide application decisions are typically made on a per-tree basis. When over 15% of detected shoots require treatment, the tree is designated for spraying;Validation against manual assessments yielded the following model performance metrics: detection rate 89.10%, miss rate 7.35%, false detection rate 3.55%, with pesticide decision accuracy reaching 83.3%. (Detailed data are listed in Table 4, and among the six tested trees, the lengths of winter shoots were divided into three categories (short, medium, and long), with two trees per category. The corresponding detection results are presented in Figure 12). The experimental results conclusively demonstrate the model's outstanding performance in lychee winter shoot identification tasks: not only achieving an overall detection accuracy approaching 90%, but also maintaining high reliability in pesticide application decisions. Particularly noteworthy is the model's consistently stable recognition performance and exceptional generalization capability when handling diverse sample detection requirements.

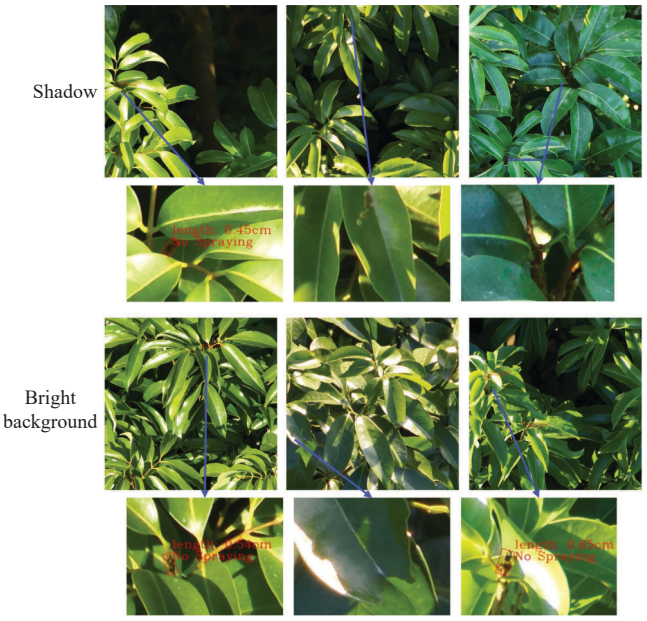


Figure 11 YOLOv8n_OBB_SEB inference results

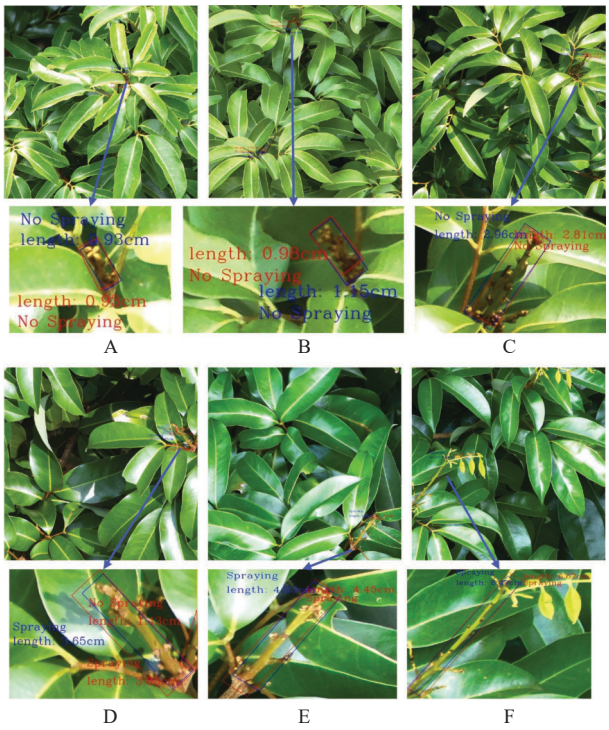
Table 4 Experimental results of the YOLOv8n_OBB_SEB model on six lychee trees

Tree	Detected	Undetected	Real	False	Detection rate/%	Undetection rate/%	False rate/%	Actual drug requirement	Estimated drug requirement
A	31	2	34	1	91.2	5.9	2.9	No	No
B	37	3	40	0	92.5	7.5	0	No	No
C	30	3	33	0	90.9	9.1	0	No	Yes
D	11	0	13	2	84.6	0	15.4	Yes	Yes
E	14	2	16	0	87.5	12.5	0	Yes	Yes
F	29	3	33	1	87.9	9.1	3.0	Yes	Yes

4 Discussion and conclusions

4.1 Discussion

Although the proposed YOLOv8n_OBB_SEB model shows significant improvements across multiple evaluation metrics, the experimental results also reveal several issues that need to be addressed. Firstly, some lychee winter shoots experienced false positives or missed detections, indicating that the model's performance in complex backgrounds still has room for improvement, which is a common challenge in small object detection. Secondly, some rotated bounding boxes did not accurately fit the actual boundaries of the targets, leading to errors in measuring the target's length. To address these issues, future research could focus on the following aspects: first, introducing more diverse training datasets, covering different lighting conditions, complex backgrounds, and varying target shapes, to enhance the model's robustness and adaptability, enabling it to better handle the challenges of real-world scenarios; second, optimizing the model's loss function design, particularly improving the penalty mechanism for target box localization errors, to increase the model's sensitivity to differences between predicted and actual boxes, thereby improving the precision of the target boxes; third, integrating multi-scale feature extraction techniques to further enhance the model's ability to capture detailed features and reduce



Note: Red: model prediction; blue: ground truth.

Figure 12 Detection results of YOLOv8n_OBB_SEB model on six trees (A, B, C, D, E, F)

the occurrence of false positives and missed detections. At the same time, using rotated bounding boxes to estimate the length of lychee winter shoots may still introduce certain errors, primarily in the following aspects:

1) The limitation of laser rangefinders lies in that such ranging devices can only measure the distance at the focal point, which leads to measurement deviations for other points in the image. This deviation is particularly prominent when the density of lychee canopies is low, as it causes significant increases in the distance differences between different pixels in the image and the camera. To reduce this error, improvements can be made by equipping drones with LiDAR, thereby minimizing the error.

2) Fit between the detection box and the winter shoot: The rotated detection box may not perfectly align with the boundaries of the lychee winter shoot, leading to measurement deviations. This issue can be addressed by further optimizing the model's detection accuracy and boundary-fitting capability. Through detection and calculation, the four corner coordinates of the target's rotated bounding box are obtained. After computing the pixel length and converting it to the actual dimension, the mean absolute error (MAE) on the test set is 7.3%.

3) Curvature of the lychee winter shoot: Estimating the length of the lychee winter shoot using the length of the rotated bounding box introduces an error, as the winter shoots may be curved. However, observations of previously captured images show that newly sprouted lychee shoots are relatively straight. It is only after a certain period of growth that the shoots begin to bend slightly. By then, most of the winter shoots will have reached the stage where pesticide spraying is required, so this error is considered acceptable.

In future work, we plan to integrate the YOLOv8n_OBB_SEB model into a smart orchard platform, continue data collection, and use the model for inference. By incorporating feedback from detection performance, we will continuously optimize the model to better suit complex and diverse environments, meeting the needs of agricultural production.

4.2 Conclusions

To achieve accurate detection of lychee winter shoots in real orchard environments, this study proposes a method combining drone remote sensing technology with the YOLOv8n_OBB_SEB model for classifying the lengths of winter shoots. The model introduces the StarBlock module into the backbone network and embeds the C2f module to increase network dimensionality and enhance feature extraction capabilities, while replacing the traditional Concat module with BiFPN, dynamically adjusting the importance of features through a weighting mechanism to achieve bidirectional feature fusion and further optimize the fusion process. In addition, it integrates the EMA attention module into the C2f module to help the model more accurately distinguish between targets and backgrounds, thereby improving detection accuracy. Experimental results show that on the self-constructed lychee winter shoot dataset, the YOLOv8n_OBB_SEB model achieves an average precision (AP) of 89.2%, a 20.7% improvement over the original YOLOv8n_OBB model. To verify the model's generalization performance, the tests at a flight height of 7 m above the trees were also conducted, achieving a precision of 88.7%, and comparative experiments with other lychee winter shoot detection models further confirm that this model can achieve higher detection accuracy in complex backgrounds. Meanwhile, considering the practical needs of lychee production, merely detecting lychee winter shoots is insufficient to directly solve production issues, so rotated bounding boxes were used to fit the shape of lychee winter shoots and estimate their lengths to facilitate subsequent evaluation of whether pesticide spraying is necessary. Future research will focus on optimizing detection accuracy and the fitting degree of rotated bounding boxes, and explore the application of this method in actual orchard management to support the intelligent and precise management of agricultural production.

Acknowledgement

The author expresses gratitude for the support of the Guangdong Provincial Key R&D Program (Grant No. 2023B0202090001), the National Natural Science Foundation of China General Program (Grant No. 32371984), the National Key Research and Development Plan Project (Grant No. 2023YFD 2000200), the Guangdong Higher Education Key Area (Artificial Intelligence) Special Project (Grant No. 2019KZDZX1012), the '111 Center' (Grant No. D18019) and China Agriculture Research System (Grant No. CARS-15-22) in this study. The author would also like to thank all those who contributed to this research and the authors cited in the paper.

[References]

[1] Chen W S, Ku M L. Ethephon and kinetin reduce shoot length and increase flower bud formation in lychee. HortScience, 1988; 23(6): 1078–1078.

[2] Koul B, Taak P. Lychee (*Litchi chinensis* Sonn.): Pre- and Post-harvest disease management. In: Kumar M, Kumar V, Bhalla-Sarin N, Varma A. (Ed.) Lychee disease management. Springer, 2017; pp.1-26. doi: 10.1007/978-981-10-4247-8_1.

[3] Huang X, Zeng L, Huang H B. Lychee and longan production in China. Acta Horticulturae, 2005; 665: 27–36.

[4] Zhao Z Q, Zheng P, Xu S T, Wu X D. Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 2019; 30(11): 3212–3232.

[5] Lu J Q, Lin W Z, Chen P F, Lan Y B, Deng X L, Niu H Y, et al. Research on lightweight citrus flowering rate statistical model combined with anchor frame clustering optimization. Sensors, 2021; 21(23): 7929.

[6] Vásconez J P, Vásconez I N, Moya V, Calderón-Díaz M J, Valenzuela M,

Besoain X, et al. Deep learning-based classification of visual symptoms of bacterial wilt disease caused by *Ralstonia solanacearum* in tomato plants. Computers and Electronics in Agriculture, 2024; 227: 109617. doi: 10.1016/j. compag. 2024.109617. doi: 10.1016/j.compag.2024.109617.

[7] Mo J W, Lan Y B, Yang D Z, Wen F, Qiu H B, Chen X, et al. Deep learning-based instance segmentation method of litchi canopy from UAV-acquired images. Remote Sensing, 2021; 13(19): 3919.

[8] Huang W X, Huo Y, Yang S C, Liu M J, Li H, Zhang M. Detection of *Laodelphax striatellus* (small brown planthopper) based on improved YOLOv5. Computers and Electronics in Agriculture, 2023; 206: 107657.

[9] Li Z K, Deng X L, Lan Y B, Liu C J, Qing J J. Fruit tree canopy segmentation from UAV orthophoto maps based on a lightweight improved U-Net. Computers and Electronics in Agriculture, 2024; 217: 108538.

[10] Song C Y, Zhang F, Li J S, Xie J Y, Yang C, Zhou H, et al. Detection of maize tassels for UAV remote sensing image with an improved YOLOX model. Journal of Integrative Agriculture, 2023; 22(6): 1671–1683.

[11] Qing J J, Deng X L, Lan Y B, Li Z K. GPT-aided diagnosis on agricultural image based on a new light YOLOPC. Computers and Electronics in Agriculture, 2023; 213: 108168.

[12] Liang J T, Chen X, Liang C J, Long T, Tang X Y, Shi Z M, et al. A detection approach for late-autumn shoots of litchi based on unmanned aerial vehicle (UAV) remote sensing. Computers and Electronics in Agriculture, 2023; 204: 107535.

[13] Kisantal M, Wojna Z, Murawski J, Naruniec J, Cho K. Augmentation for small object detection. arXiv, 2019; In press. Available from: http://arxiv.org/abs/1902.07296. Accessed on [2025-05-05].

[14] Noh J, Bae W, Lee W, Seo J, Kim G. Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, 2019; pp.9725–9733. doi: 10.1109/ICCV.2019.00982.

[15] Akyon F C, Altinuc S O, Temizel A. Slicing aided hyper inference and fine-tuning for small object detection. In: 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, 2022; pp.966–970. doi: 10.1109/ICIP46576.2022.9897990.

[16] Zhu X K, Lyu S C, Wang X, Zhao Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, 2021; pp. 2778–2788. doi: 10.1109/ICCVW54120.2021.00312.

[17] Cao Y R, He Z J, Wang L J, Wang W G, Yuan Y X, Zhang D W, et al. VisDrone-DET2021: The vision meets drone object detection challenge results. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, 2021; pp.2847–2854. doi: 10.1109/ICCVW54120.2021.00319.

[18] Cai X H, Lai Q X, Wang Y W, Wang W G, Sun Z R, Yao Y Z. Poly kernel inception network for remote sensing detection. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 2024; pp.27706–27716. doi: 10.1109/CVPR52733.2024.02617.

[19] Lim J S, Astrid M, Yoon H J, Lee S I. Small object detection using context and attention. In: 2021 IEEE International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Jeju Island, 2021; pp.181–186. doi: 10.1109/ICAIIC51459.2021.9415217.

[20] Liu Q, Lyu J, Zhang C P. MAE-YOLOv8-based small object detection of green crisp plum in real complex orchard environments. Computers and Electronics in Agriculture, 2024; 226: 109458.

[21] Ma X, Dai X Y, Bai Y, Wang Y Z, Fu Y. Rewrite the Stars. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024; pp.5694–5703. doi: 10.1109/CVPR52733.2024.00544.

[22] Tan M X, Pang R M, Le Q V. EfficientDet: Scalable and efficient object detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle: IEEE, pp.10778–10787. doi: 10.1109/CVPR42600.2020.10179.

[23] Ouyang D L, He S, Zhang G Z, Luo M Z, Guo H Y, Zhan J, et al. Efficient multi-scale attention module with cross-spatial learning. In: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island: IEEE, 2023; pp.1–5. doi: 10.1109/ICASSP49357.2023.10096516.

[24] Lin T Y, Dollar P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, pp.936–944. doi: 10.1109/CVPR.2017.106.

[25] Chen J, Mai H S, Luo L B, Chen X Q, Wu K L. Effective feature fusion network in BIFPN for small object detection. In: 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, 2021; pp.699–703. doi: 10.1109/ICIP42928.2021.9506347.

[26] Mo H H, Wei L J. Tomato yellow leaf curl virus detection based on cross-domain shared attention and enhanced BiFPN. Ecological Informatics, 2025; 85: 102912.

[27] Zhang Z Y. Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, 1999; pp.666–673. doi: 10.1109/ICCV.1999.791289.

[28] Woo S, Park J, Lee J Y, Kweon I S. CBAM: Convolutional block attention module. In: Computer Vision - ECCV 2018, Springer, 2018; 11211: 3–19.

[29] Lau K W, Po L M, Rehman Y A U. Large separable kernel attention: Rethinking the large kernel attention design in CNN. Expert Systems with Applications, 2024; 236: 121352.

[30] Chen G, Wang H, T Chen K, Li Z J, Song Z D, Liu Y L, et al. A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2022; 52(2): 936–953. doi: 10.1109/TSMC.2020.3005231.

[31] Tong K, Wu Y Q. Deep learning-based detection from the perspective of small or tiny objects: A survey. Image and Vision Computing, 2022; 123: 104471.

[32] Xie X X, Cheng G, Wang J B, Yao X W, Han J W. Oriented R-CNN for object detection. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, 2021; pp.3500–3509. doi: 10.1109/ICCV48922.2021.00350.

[33] Yang X, Yan J, Feng Z, He T. R3Det: Refined single-stage detector with feature refinement for rotating object. Proceedings of the AAAI Conference on Artificial Intelligence, 2021; 35(4): 3163–3171.

[34] Han J, Ding J, Li J, Xia G S. Align deep features for oriented object detection. IEEE Transactions on Geoscience and Remote Sensing, 2022; 60: 1–11.

[35] Han J M, Ding J, Xue N, Xia G S. ReDet: A rotation-equivariant detector for aerial object detection. In: 2021 IEEE/CVF Conference on Coputer Vision and Pattern Recognition (CVPR), Nashville, 2021; pp.2785–2794. doi: 10.1109/CVPR46437.2021.00281.