Strawberry Fruit Recognition and Picking Point Localization Based on Lightweight YOLOv8

LIU Yizhe[1], WANG Wei[1*], Seung-Chul YOON[2], Xinzhi NI[3], GAO Ning[1], WANG Xiaorong[1], WEI Chaojie[1], YE Jiawei[1]

(1.Beijing Key Laboratory of Optimization Design for Modern Agricultural Equipment, College of Engineering, China Agricultural University, Beijing 100083, China)

(2.U. S. National Poultry Research Center, USDA-ARS, 950 College Station Rd., Athens, GA 30605, USA)

(3.Crop Genetics and Breeding Research Unit, USDA-ARS, 2747 Davis Road, Tifton, GA 31793, USA)

Abstract: The aim of this study was to address the issues of inaccurate picking point localization caused by overlapping and occlusion of ridge-planted strawberry fruits, as well as storage quality deterioration induced by postharvest perishability. A lightweight object detection model was developed to rapidly and precisely localize strawberry fruits and picking points, thereby providing technical support for intelligent harvesting and postharvest supply chain quality control. An improved YOLOv8-SCE model was developed by replacing the C2f module in the YOLOv8 backbone with the C2f-fasternet module to reduce computational redundancy, introducing an Efficient Multi-Scale Attention (EMA) mechanism to enhance multi-scale feature representation and adopting the SIoU loss function for bounding-box regression optimization. Additionally, a novel picking point prediction algorithm was proposed based on fruit-pedicel spatial relationships. The experimental results demonstrate that the improved model achieved a precision (P), recall (R), and mean average precision (mAP@0.5) of 99.23%, 98.11%, and 99.02%, respectively, on the test set, with a 25.22% reduction in parameter count and an 8.93 FPS improvement in inference speed. The accuracy of picking point prediction reached 91.11%, with single-fruit prediction time of only 96 ms. Practical deployment tests show that both recognition and localization algorithms maintained excellent performance, achieving accuracy rates above 92.07%. Compared with traditional machine vision methods and mainstream deep learning models, the proposed method exhibited superior precision and enhanced time efficiency. These findings confirm that the YOLOv8-SCE model and its associated picking point prediction method provide a solid theoretical foundation and technological innovation pathway for achieving intelligent, precise harvesting operations in strawberry-harvesting robots.

Key words: lightweight YOLOv8; strawberry fruits and pedicels recognition; picking point; attention mechanism

Article No.: 1673-9078(2026)7-272-283 DOI: 10.13982/j.mfst.1673-9078.2026.7.0669

基于轻量化YOLOv8的草莓果实识别与采摘点定位

刘义哲[1]，王伟[1*]，Seung Chul YOON[2]，Xinzhi NI[3]，高宁[1]，王笑荣[1]，魏超杰[1]，叶嘉伟[1]

（1.中国农业大学工学院，现代农业装备优化设计北京市重点实验室，北京 100083）

（2.美国农业部农业研究服务局，美国家禽研究中心，佐治亚州雅典市学院站路950号 30605）

（3.美国农业部农业研究服务局，作物遗传与育种研究组，佐治亚州蒂夫顿市戴维斯路2747号 31793）

引文格式：

刘义哲,王伟,Seung Chul YOON,等.基于轻量化YOLOv8的草莓果实识别与采摘点定位[J].现代食品科技,2026,42(7): 272-283.

LIU Yizhe, WANG Wei, Seung-Chul YOON, et al. Strawberry fruit recognition and picking point localization based on lightweight YOLOv8 [J]. Modern Food Science and Technology, 2026, 42(7): 272-283.

Submission date: 2025-05-14; Revision date: 2025-07-21; Acceptance date: 2025-07-28

Acknowledgments: National Natural Science Foundation of China (32272410); China Scholarship Council (202306350090)

First authors: LIU Yizhe (2000-), male, Master's candidate, research direction: Information and Image Processing Technology, E-mail: S20223071380@cau.edu.cn

Corresponding author: WANG Wei (1975-), male, Ph.D., Professor, research direction: Information and Image Processing Technology, E-mail: playerwxw@cau.edu.cn

摘要：为应对垄植草莓果实堆叠、遮挡导致的采摘点定位不准及采后易腐性引发的贮藏品质劣变问题，本研究旨在开发一种轻量化目标检测模型，实现草莓果实与采摘点的快速精准定位，为智能化采收及采后产业链质量控制提供技术支持。通过将 YOLOv8 主干网络 C2f 模块替换为 C2f-fasternet 模块降低计算冗余，引入多尺度注意力机制（Efficient Multi-Scale Attention, EMA）强化多尺度特征表征，采用 SIoU 损失函数优化检测框回归，形成改进的 YOLOv8-SCE 模型；同时基于果实－果梗空间关系提出新型采摘点预测算法。该研究结果为改进模型在测试集上的精确率（P）、召回率（R）以及平均精度（mAP@0.5）分别达 99.23%、98.11% 以及 99.02%，参数量减少 25.22% 且推理速度提升 8.93 FPS，采摘点预测准确率 91.11% 且单果预测耗时仅 96 ms，实际部署测试下识别与定位算法准均有着良好的表现，准确率均在 92.07% 以上。较传统机器视觉方法及主流深度学习模型兼具更高精度与更优时效性。该研究结论为本研究提出的 YOLOv8-SCE 模型及其采摘点预测方法为草莓采摘机器人实现智能化、精准化采收作业提供了坚实的理论基础与技术创新路径。

关键词：轻量级 YOLOv8；草莓果实与果梗识别；采摘点；注意力机制

Strawberries, as a beloved fruit among consumers, have a stable market demand and are extensively cultivated worldwide. During the harvest season of strawberries, fruits are typically harvested when approximately three-quarters of their surface has turned red. However, due to the tendency of the fruits to become excessively ripe, strawberries can turn yellow, moldy, or even rotten, making it crucial to strictly control the harvest time. Due to soft texture of strawberries, current strawberry harvest primarily relies on manual labor, which involves high labor intensity and low harvesting efficiency[1]. Therefore, improving harvest efficiency, reducing labor costs, and promoting the mechanization of strawberry harvest have become one of the urgent issues to address in the field of agricultural engineering research[2,3].

More and more researchers are dedicated to assessing the key technologies of strawberry harvest robots. Including the recognition of ripe strawberry fruits, precise localization and picking the strawberry fruits, and model deployment are being identified as three crucial components[4,5].

Currently, research has made progresses in the recognition of strawberry fruits and localization the picking points. Traditional methods involve using image recognition technology to extract information such as the shape, size, color, and defects of strawberries for detection. However, those methods were susceptible to environmental factors and often had relatively low accuracy[6,7]. With the widespread application of machine vision technology in agricultural harvest in general, the accuracy of fruit detection has significantly improved[8,9]. Recent developments include using deep learning for classifying ripe and raw strawberries, with advancements in semantic segmentation models achieving high accuracies for mature and immature fruit detection. Despite improvements, challenges such as background interference and processing speeds in real-world environments remain critical areas of focus for further research[10-12].

Compared to previous algorithms, the YOLO series' one-stage detection algorithms have shown superiority in addressing agricultural detection challenges. These include RTSD-Net based on YOLOv4-tiny for real-time indoor strawberry detection using embedded systems like Jetson Nano[13]. However, existing experimental datasets often lack diverse occlusion scenarios, limiting their practical applicability in natural or field conditions. Recent studies such as YOLOv5s-BiCE and YOLOv5-ASFF have improved detection accuracy and generalization capabilities but have increased computational demands[14,15]. Overall, the YOLO series represents a promising approach for fruit recognition, yet ongoing efforts are needed to optimize detection speed while maintaining high accuracy levels[16,17].

In current research, various methods have been explored for determining picking points on strawberry pedicels. Approaches like Mask-RCNN faced challenges with occlusions and detection speed, limiting practical field use[18]. Recent advancements included perception systems utilizing point cloud data for vertical environments, but issues with segmentation errors persisted in clustered strawberries[19]. An improved semantic segmentation model for tomatoes showed promise, achieving a high accuracy rate in picking point localization[20], providing a theoretical basis for picking points localization in the study.

Although some progress has been made, current technologies still struggle to effectively apply strawberry recognition and precise picking points localization of strawberry fruits in the practical operations of strawberry harvest robots. In the environment of raised-planting cultivation, strawberry harvest robots are required to have high computational capabilities, accurate recognition of strawberry fruits, and precise localization of the picking points. Additionally, raised-planting cultivation is closer to the natural growth habits of strawberries, making research on ridge-planted strawberries more practically valuable. Therefore, the current study focuses on ridge-planted strawberries, by designing a lightweight YOLOv8-SCE model, which significantly reduces the number of model parameters while ensuring accurate recognition of strawberry fruits and precise position picking points, making it suitable for deployment on mobile platforms.

1 Materials and Methods

1.1 Dataset Construction

Strawberry images for the study were captured at a commercial strawberry farm in Beijing, China, where ridge-planted strawberries were grown in a greenhouse. The weather was clear with full sunlight outside the greenhouse during the image acquisition session in the afternoon of January 28, 2024. A handheld smartphone with a resolution of 3 072 (width)×2 304 (height) pixels was used to collect the JPEG images of the strawberry trees under natural growing conditions with varying degrees of fruit ripeness. A total of 480 images were selected for this study, encompassing various lighting conditions and different occlusion scenarios observed at a working distance of 10~30 cm. This dataset was designed to reflect the challenges faced by strawberry picking robots in real-world environments, with varying lighting, occlusion, and fruit ripeness conditions. The 10~30 cm working distance during image capture aligns with the typical operational range of a strawberry picking robot's vision system. Thus, this dataset is highly relevant for training and evaluating robotic vision models for strawberry harvesting.

To enhance the generality and robustness of the detection model in various complex real-world scenarios, the captured dataset of 480 images was expanded to 1 920 images by data augmentation. The data augmentation included brightness adjustments to simulate different light intensities, adding noise to simulate blurring caused by camera shaking during handheld image acquisition, and random rotation (clockwise between −45° and 45°) to simulate different shooting angles. After data augmentation, all augmented versions of the same original image were assigned to the same subset (training, validation, or testing), ensuring no overlap between the subsets. Labeling of the targets was performed using LabelImg (an open-source image annotation software) for three categories: ripe strawberry fruits, raw strawberry fruits, and pedicels. Strawberry fruits with over 80% color redness were labeled as "ripe", the remaining unripe strawberry fruits were labeled as "raw", and the pedicel above the strawberry fruit was labeled as "pedicel". The annotation results were stored in XML format, including image path, pixel information, target categories, and position information of the annotation boxes. For ease of training, the annotation files were batch converted to txt format using Python. Finally, the annotated dataset was divided into training, validation, and testing in a 7:2:1 ratio, with 1 344 images for training, 384 images for validation, and 192 images for testing, as shown in Table 1.

Table 1 Dataset classification

Dataset	Number of images	Number of tags		
		ripe	raw	pedicel
Training set	1 344	2 967	3 314	7 295
Verification set	384	985	806	2 059
Test set	192	423	473	1 042
total	1 920	4 375	4 593	10 396

1.2 Proposed Strawberry Object Detection Model

Based on the advantages of YOLOv8 in terms of speed, accuracy, and flexibility[21,22], this study chose YOLOv8 as the base model for target detection in ridge-planting strawberry fields. Compared to earlier versions like YOLOv4 and YOLOv5, YOLOv8 strikes a perfect balance between high precision and real-time performance, which is crucial for tasks requiring both accuracy and rapid response, like strawberry harvesting in dynamic environments.

YOLOv8 structure is divided into three main components: the backbone, neck, and head. The backbone consists of essential convolutional modules (Conv), spatial pyramid pooling module (SPPF), and C2f module, responsible for feature extraction. The neck includes the PAN (Path Aggregation Network) and FPN (Feature

Pyramid Network) structures, forming the PANet architecture, effectively enhancing the network's ability to fuse features from objects at different scales. The head serves as the classifier and regressor, utilizing a decoupled head structure that separates classification and detection tasks, allowing focused improvement on each task and enhancing overall performance.

This study improves the YOLOv8 model while maintaining high recognition accuracy, resulting in the lightweight YOLOv8-SCE model. The improvements included replacing the C2f module in the YOLOv8 model with the C2f-faster module to reduce redundant computations and memory access, thereby enhancing the efficiency of spatial feature extraction. Building upon C2f-faster, the EMA attention mechanism was introduced, resulting in C2f-faster-EMA to further enhance model performance. Finally, the SIoU loss function was utilized to achieve rapid localization of predicted bounding boxes. The improved network structure was shown in Fig.1.

1.2.1 FasterNet Block

To achieve the recognition task of strawberry fruits and pedicels under various complex conditions, convolutional neural networks were needed to perform a large number of filtering operations and feature extraction, which not only increased the complexity of the model but also required more computational resources for training and inference.

To reduce the model's complexity and improve network processing speed, this study proposed the C2f-faster module, it was derived by replacing the BottleNeck in C2f with the Fasternet Block[23], which resulting in enhanced feature extraction efficiency (Fig.2a). The Fasternet Block introduced a novel partial convolution PConv, which only applied regular convolution for spatial feature extraction on a part of the input channels and keeps the remaining channels unchanged, as shown in Fig.2b. This module is highly effective in processing images with obstructions or partial missing areas, significantly reducing memory access overhead while maintaining network performance.

1.2.2 EMA (efficient multi-scale attention)

Additionally, in practical environments, factors such as fruits overlapping, pedicels intertwining, and complex backgrounds can affect the recognition accuracy of images. Therefore, in order to further improve the efficiency and performance of the model, focusing neural networks on extracting features of strawberry fruits and pedicels while disregarding other information in the image, this study incorporated the use of attention mechanisms to enhance the model's acquisition of critical information from the images.

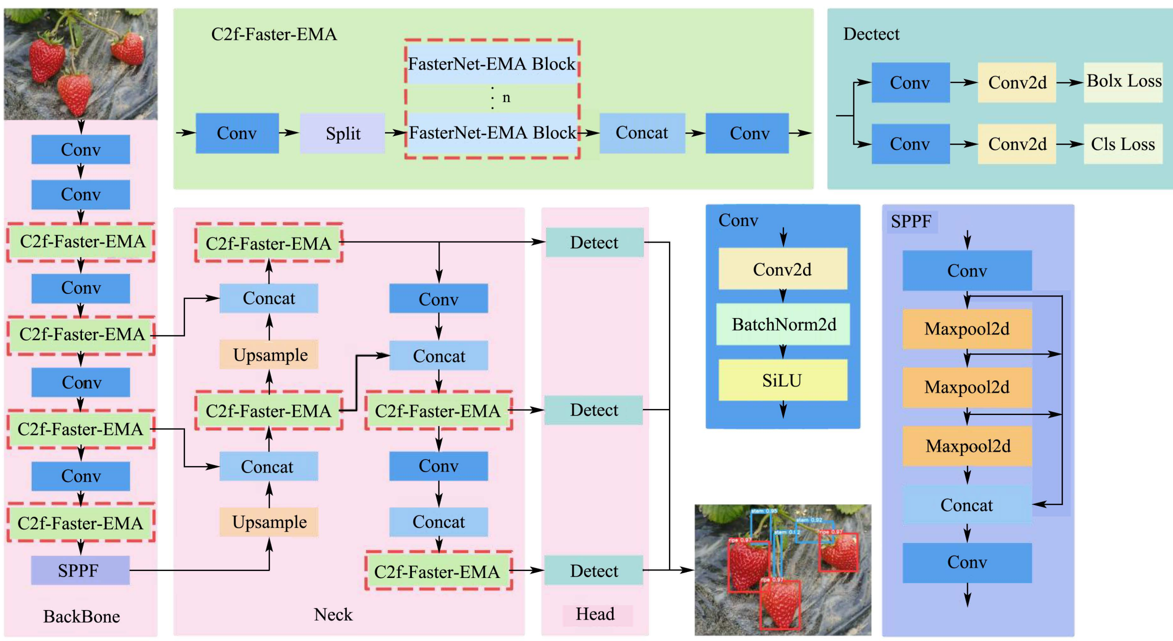

Fig.1 The architecture of the improved model

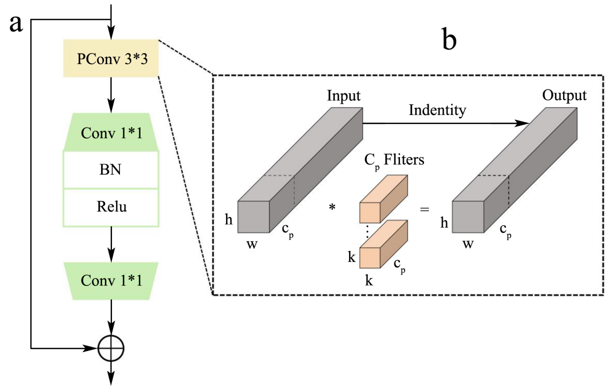

Fig.2 FasterNet Block and its key components

Notes: (a) FasterNet Block; (b) Partial convolution.

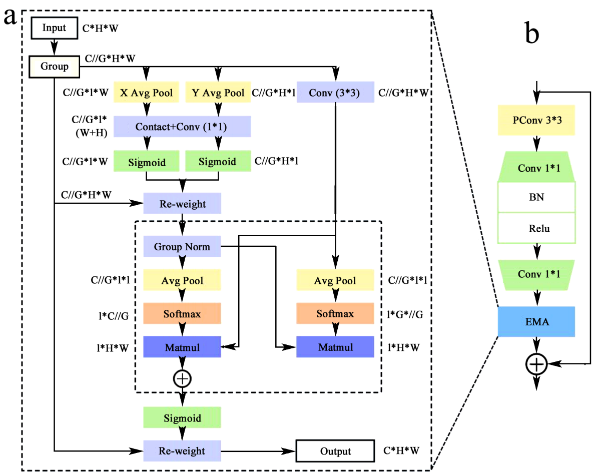

Fig.3 FasterNet-EMA Block and its key components

Notes: (a) EMA; (b) FasterNet-EMA Block.

Attention mechanisms improve recognition accuracy and scene understanding in computer vision tasks[24,25], but they require significant memory and computational resources. An efficient multi-scale attention (EMA) module based on the CA module has been utilized in recent studies[26], which not only enhances performance but also reduces computational overhead, as shown in Fig.3a. By reshaping part of the channel into the batch dimension and grouping the channel dimension into multiple sub-features, EMA ensures the uniform distribution of spatial semantic features in each feature group. This helps to effectively establish short-distance and long-distance dependencies for better performance. Leveraging the flexibility and lightweight nature of EMA, we integrated it into the forward propagation process of FasterNet, resulting in FasterNet-EMA, as shown in Fig.3b.

1.2.3 SIoU

YOLOv8 defaults to using CIoU as the bounding box regression loss function, evaluating the difference in aspect ratio between the predicted box and the target box to assess the positional differences between them. However, the complete intersection over union (CIoU) does not effectively handle aspect ratio and angle constraints between predicted and ground truth boxes[27]. Therefore, this study introduced SIoU as the detection box regression loss function[28]. The SIoU loss function incorporates the orientation between the ground truth box and the predicted box, introducing angle vectors as constraints to improve network convergence and training quality, it consists of four cost functions: angle loss, distance loss, shape loss, and IoU loss, the calculation formulas are shown in Equations 1~4.

$$L_{SIOU}=1-IOU+\frac{\Delta+\Omega}{2} \tag{1}$$

$$IOU=\frac{|B\cap B^{GT}|}{|B\cup B^{GT}|} \tag{2}$$

$$\Lambda=1-2\sin^2\left[arc\sin\left(\frac{C_h}{\sigma}\right)-\frac{\pi}{4}\right] \tag{3}$$

$$\Omega=\sum_{t=w,h}(1-e^{-\omega_t})^{\theta} \tag{4}$$

In the formula:

IOU——represents the IOU loss,

Λ——represents the angle loss,

Δ——represents the distance loss considering the angle loss condition,

Ω——represents the shape loss,

B and B^{GT}——represent the areas of the predicted bounding box and the ground truth bounding box respectively.

1.3 Proposed Picking Point Position Method

After obtaining images of ripe strawberry fruits using YOLOv8-SCE, the edge contours are extracted through threshold segmentation to determine the apex of the fruit. Subsequently, calculate the distance between the midpoint of the bottom edge of all pedicel bounding boxes and the apex of the fruit, select the pedicel closest to the ripe strawberry fruits, and thus position the picking points.

First, calculated the Euclidean distance between the lowest point of each pedicel's bounding box and the apex of the fruit to filter out the pedicel closest to a fruit.

Next, to avoid mismatching irrelevant pedicels, further determination was conducted to ascertain whether this distance is less than 1/4 of the fruit's height. If so, the pedicel was confirmed to belong to the fruit.

(a) Estimation of the apex of the fruit. First, cropped the detected bounding box area to generate a sub-image. Next, extracted the region of interest in the HSI color space to clearly identify the fruit's contour and address noise and holes introduced in step (1) through necessary image processing such as noise removal and hole filling. Subsequently, the minimum extreme point of the strawberry contour was determined as the lowest point of the fruit. Extending the line connecting the center of the bounding box with this point, and took the intersection with the bounding box as the highest point of the extension.

(b) Determining the pedicel to which the fruit belonged. First, the Euclidean distance between the lowest point of each pedicel's bounding box and the apex of the fruit was calculated to filter out the pedicel closest to the fruit. Next, to avoid mismatching irrelevant pedicels, this distance was further determined whether it is less than 1/4 of the fruit's height or not. If so, the pedicel was confirmed to belong to the fruit.

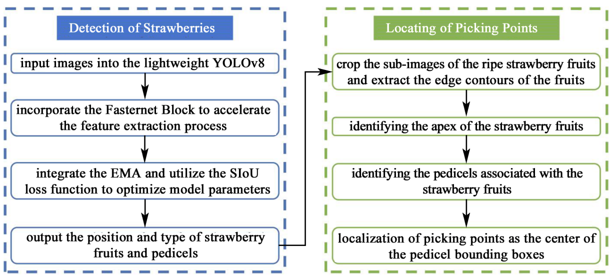

Fig.4 Flowchart of methodology

(c) localization the picking points. To reduce collisions during the harvest, the picking points should be set approximately 5~20 millimeters above the apex of the fruit. Experiments have shown that about 95% of the distances between the center point of the pedicel's bounding box and the apex of the fruits are within this range. Therefore, the center point of the pedicel's bounding box was approximated as the harvest target fruit. The method flowchart of the recognition of strawberry fruits and picking points localization is shown in Fig.4.

1.4 Training Environment and Evaluation Metrics

The system environment of this experiment was a Windows 64-bit system, running on a server equipped with an Intel (R) Core(TM)i5-10 400F CPU @2.90GHz 2.90GHz processor, with 16GB of RAM, and using NVIDIA GeForce GTX 1 650 graphics card. The software includes CUDA11.1 and CUDNN11.1. Programming was conducted using Python, with model building and training carried out in the PyTorch framework. To accelerate the convergence speed during model training, the SGD optimizer was used, with 300 training iterations and a batch size of 16. In the training stage of the network model, all input images were resized to 640×640, the initial learning rate was set to 1×10^{-3}, momentum was set to 0.94, and a cosine annealing decay strategy was used to update the learning rate.

To evaluate the accuracy of the improved network model in recognizing strawberry fruits and pedicels, five evaluation metrics are adopted: Precision (P), Recall (R), Mean Average Precision (mAP@0.5), and F1-score. When calculating mAP@0.5, since the focus of the current study was on recognizing strawberry pedicels, ripe fruits, and unripe fruits, the number of classes were three. The calculation formulas for the relevant performance metrics are shown in Eq. (5)~(10). Number of parameters and Giga Floating-point Operations Per Second (GFLOPs) were used to measure the complexity of the model. The detection speed of the model is measured in Frames Per Second (FPS). A high FPS means the model can process data quickly and efficiently, meeting the real-time requirements of practical applications.

$$Precision=\frac{TP}{TP+FP} \tag{5}$$

$$Recall=\frac{TP}{TP+FN} \tag{6}$$

$$AP=\int_0^1 P(R)dR \tag{7}$$

$$mAP=\frac{\sum_{i=1}^{n} AP_i}{n} \tag{8}$$

$$F1=\frac{2\times P\times R}{P+R} \tag{9}$$

$$FPS=\frac{1}{Processing\ time\ per\ frame} \tag{10}$$

In the formula:

TP——represents the number of samples correctly identified as positive among the true positive samples;

FN——represents the number of samples incorrectly identified as negative among the true positive samples;

FP——represents the number of samples incorrectly identified as positive among the true negative samples;

TN——represents the number of samples correctly identified as negative among the true negative samples.

In Eq.9, n typically refers to the number of categories involved in the evaluation, since the focus of this paper is on identifying strawberry pedicels, ripe fruits, and raw fruits, the number of categories is n=3.

In the process of detecting picking points of ripe strawberries, accuracy is used as a metric to evaluate algorithm performance (Eq.11).

$$Accuracy=\frac{T}{N} \tag{11}$$

In the formula:

T——represents the number of correctly predicted samples;

N——denotes the total number of samples.

2 Results and Discussion

2.1 Strawberry Fruits and Pedicels Recognition Experiment

2.1.1 Ablation Experiments

This study has made three significant improvements to the baseline YOLOv8s model: optimizing the bounding box regression loss function to SIoU, replacing the C2f module with the C2f-Fasternet module, and embedding the EMA attention mechanism within the C2f-Fasternet module. The effectiveness of the enhancements was verified through ablation studies. All models were trained in a consistent hardware environment and with identical parameter settings.

As shown in Table 2, optimizing the bounding box regression loss function to SIoU enhanced the model's sensitivity to targets of varying sizes, improving both mAP@0.5 and FPS. Building on this, replacing the C2f module with the C2f-Faster module, the model's parameters and GFLOPs were reduced by 25.22% and 24.60%, significantly lowering computational requirements. This alleviates storage and processing pressure on edge devices, improving deployment efficiency. The reduction not only helps reduce latency and enhances real-time inference capabilities but also makes the model more suitable for resource-constrained devices, improving device response speed and reducing power consumption. As a result, the model's complexity is significantly improved, while recognition speed increases by 59.72%. Furthermore, integrating the EMA attention mechanism into the base model improved P from 98.73% to 99.23%, while FPS decreased from 47.63 to 38.44. However, in practical strawberry picking tasks, 30 FPS is generally sufficient to meet operational requirements. Therefore, with FPS still above this threshold, the slight decrease is a reasonable trade-off for the significant accuracy improvement, as higher precision is essential for reliable strawberry harvesting. Overall, the improved model maintained excellent recognition accuracy while it has significantly reduced the number of model parameters, which achieved the goal of lightweight model.

2.1.2 Strawberry Fruits and Pedicels Recognition Experiment

In this experiment, three mainstream one-stage detection algorithms, YOLOv5s, YOLOv7, and YOLOv8, were selected as comparative models. The experimental results are shown in Table 3. The improved YOLOv8-SCE model outperformed other models in general. Compared to the baseline YOLOv8 model, the detection results still maintained high accuracy, with mAP@0.5 and P values reaching 99.02% and 99.23%, respectively. The total parameter number of the modified model was reduced from 1.11 MB to 0.83 MB, a decrease of 25.22%, which better met the requirements of lightweight model. Fig.5 illustrates the performance differences between YOLOv8-SCE and three other models.

Table 2 Comparison results of different models for ablation experiments

YOLOv8	+SIoU	+FasterNet Block	+EMA	Parameters/M	GFLOPs/G	Precision/%	mAP@0.5/%	FPS
√				1.11	28.41	99.41	99.21	29.51
√	√			1.11	28.41	99.32	99.32	29.82
√	√	√		0.83	21.42	98.73	99.03	47.63
√	√	√	√	0.83	22.00	99.23	99.04	38.44

Table 3 Performance comparison of different models

Algorithm	Parameters(M)	GFLOPs(G)	Precision/%	R/%	mAP@0.5/%	F1/%	FPS
Yolov5	0.70	16.01	88.14	86.53	89.71	87.34	54.82
Yolov7	3.72	105.12	86.52	85.83	87.82	86.14	4.74
Yolov8	1.11	28.4	99.41	98.81	99.24	99.12	29.51
Yolov8-SCE	0.83	22.0	99.23	98.11	99.02	98.62	38.44

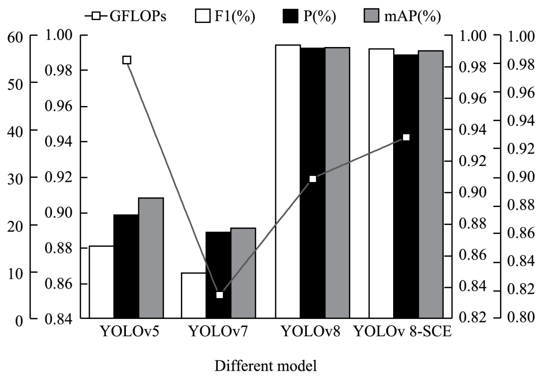

Fig.5 Performance comparison of different algorithms

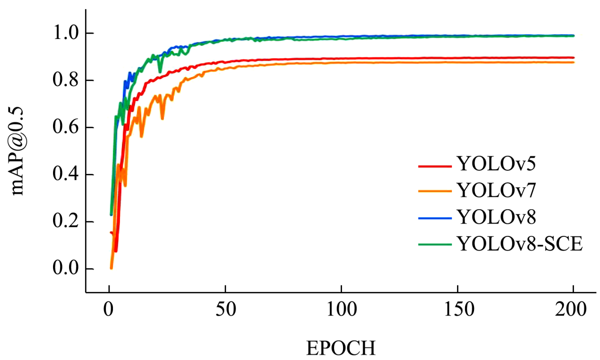

Fig.6 mAP@0.5, and iteration number variation chart

The performance levels of other models were as follows: YOLOv5 (mAP@0.5: 89.71%, P: 88.14%), YOLOv7 (mAP@0.5: 87.82%, P: 86.52%). The ranking of detection speed was YOLOv5, YOLOv8-SCE, YOLOv8, and YOLOv7. Although YOLOv5 achieved an FPS of 54.82, its mAP@0.5 and P were only 89.71% and 86.53%, indicating lower precision and not meeting the accuracy requirements for strawberry fruit and pedicel detection. It falsely identified ripe strawberry fruits as raw and failed to detect the pedicels, as shown in Fig.7a. For object detection, both YOLOv8 and YOLOv8-SCE exhibited outstanding performance. They could accurately identify the occluded strawberry fruits and pedicels in complex backgrounds without any false detections, as shown in Fig.7c and 7d, respectively. Additionally, while maintaining high-precision detection, YOLOv8-SCE processed data faster than YOLOv8, which was particularly important for applications requiring real-time fruit processing.

2.1.3 Strawberry Fruit and Pedicel Recognition under Varying Field Conditions

To validate the adaptability and robustness of the model under different field conditions, the dataset was further subdivided to facilitate assessment across diverse backgrounds and occlusion scenarios. Fig.8 shows the model's recognition results in various complex and occluded environments. In recognition tasks, the model not only accurately differentiated between ripe and raw strawberry fruits, but also precisely recognized strawberry pedicels. The method performed well under occlusions caused by other pedicels, other fruits, and leaves.

As shown in Table 4, the YOLOv8-SCE model exhibited exceptional predictive performance in recognizing strawberry fruits under various occlusion conditions, with accuracy rates exceeding 98.00% in all scenarios. Specifically, strawberry fruits occluded by pedicels achieved the highest recognition accuracy, primarily because the shape, size, and color of the pedicels minimally affected the fruit recognition. Conversely, the lowest accuracy occurred when strawberry fruits were occluded by other fruits, as the features of the occluded fruits were masked.

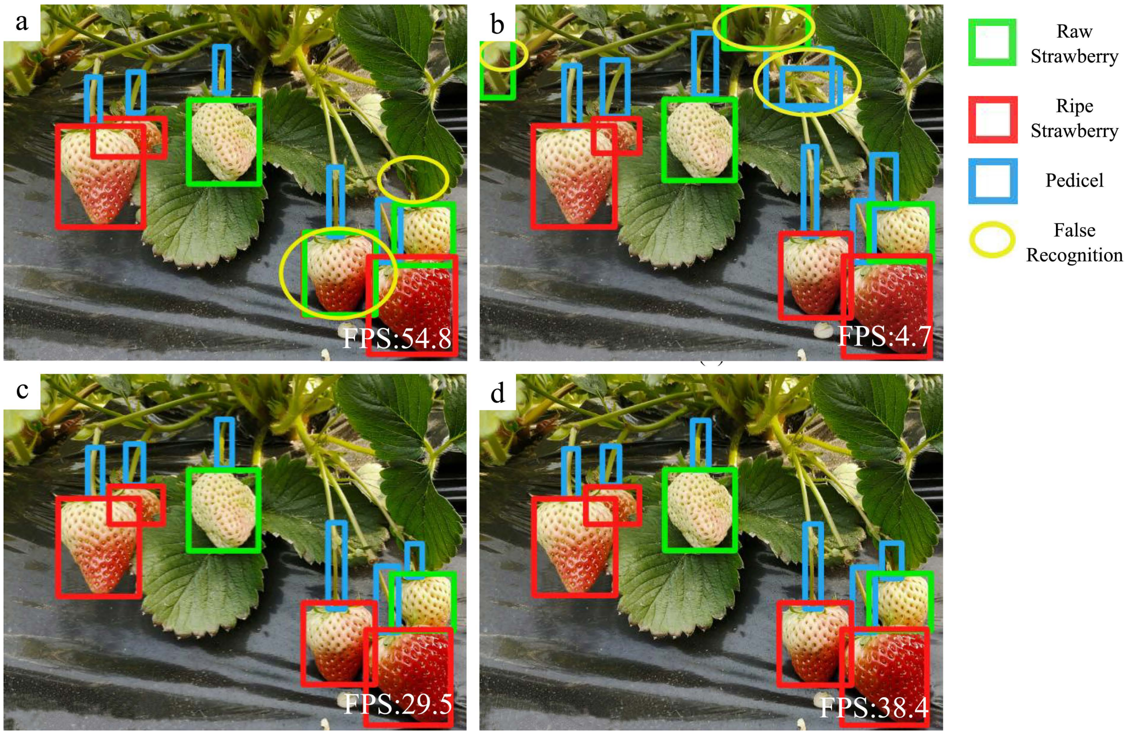

Fig.7 Comparison of recognition results of different models

Notes: (a) Yolov5; (b) Yolov7; (c) Yolov8; (d) Yolov8-SCE.

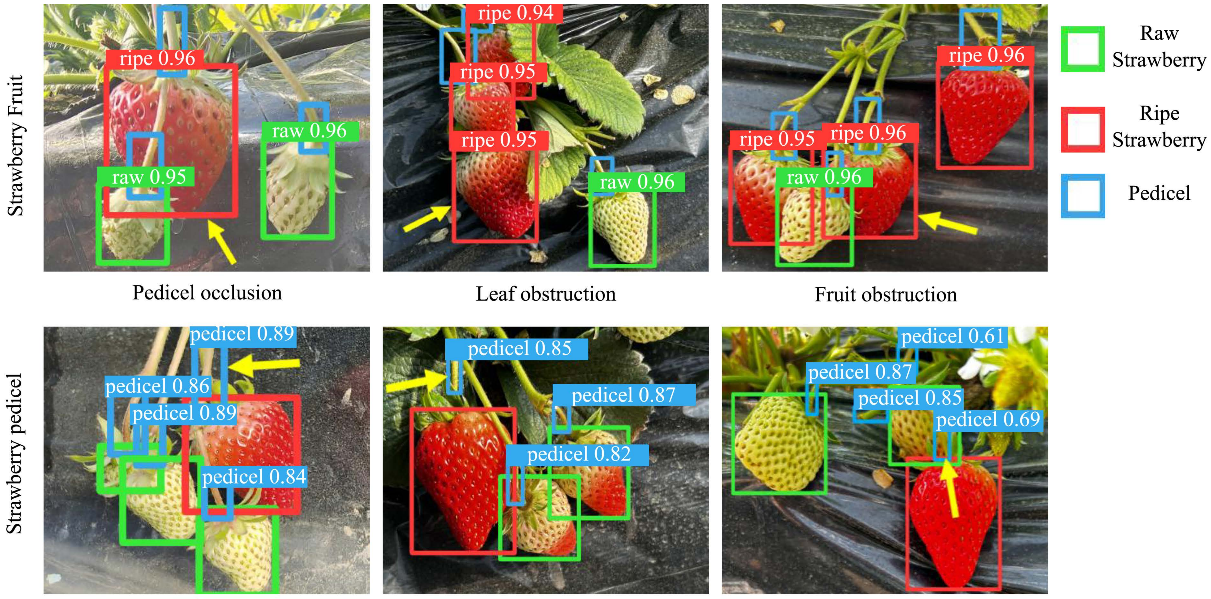

Fig.8 Detection results of strawberry fruits and pedicels

Table 4 Results of strawberry fruit and pedicel recognition experiments under different conditions

Target	Class	Detect	Correct	Correct detection rate/%
Strawberry Fruit	pedicel occlusion	338	336	99.40
	Leaf occlusion	233	230	98.71
	Fruit occlusion	195	193	98.97
Strawberry Pedicel	pedicel overlapping	253	246	97.23
	pedicel-leaf overlapping	154	152	98.70
	pedicel-fruit overlapping	100	98	98.00

In terms of pedicel detection, the YOLOv8-SCE model also demonstrated satisfying results. The accuracies in complex backgrounds of pedicel overlapping, pedicel-leaf overlapping, and pedicel-fruit overlapping were 97.23%, 98.70%, and 98.00%. The results indicated that the YOLOv8-SCE model was capable of effectively handling various complex conditions, ensuring high-precision predictions for both fruits and pedicels.

2.2 Prediction of harvest targets of ripe Strawberry Fruits

The detection process for a picking point begins with the identification of strawberry images. Using the lightweight YOLOv8-SCE model, bounding boxes were detected and positioned for strawberry fruits and pedicels, as shown in Fig.9a. Next, sub-images were cropped from the bounding boxes of each ripe strawberry fruit, and the strawberries within these sub-images underwent threshold segmentation.

The lowest point on the contour of the target strawberry fruit was marked. Then, the line connecting this point to the midpoint of the bounding box was extended. The intersection of this extended line with the bounding box was marked as the apex of the fruit (Fig.9b). Subsequently, the midpoint along the lower edge of the pedicel's bounding box was used as the detection point. With the fruit vertex as the detection center and one-quarter of the fruit height as the detection radius, the presence of an associated pedicel was identified, as shown in Fig.9c. Finally, the center position of the bounding box of the associated pedicel was identified as the picking point (Fig.9d).

The experimental results of predicting picking points for 304 ripe strawberry fruits showed accuracy was 91.11% for this method, with a target localization time of 96 ms. The incorrect localization of picking points was primarily due to the following reasons: First, YOLOv8, during the recognition of strawberry fruits and pedicels, might have produced inaccurate results due to the complex shapes of the fruits and pedicels, occlusions, or background interference, which affected subsequent judgments. Second, the threshold segmentation might have been incomplete due to uneven lighting or excessive occlusion, leading to inaccuracies in contour extraction. Lastly, when determining the pedicel associated with the strawberry fruit, the presence of closer pedicels within the distance range could have caused incorrect judgments. Although this method could not accurately position picking points from severely occluded pedicels, in practical harvest under the field conditions, these strawberry fruits were often unable to be successfully harvested, so excessive attention to errors caused by obstruction is unnecessary.

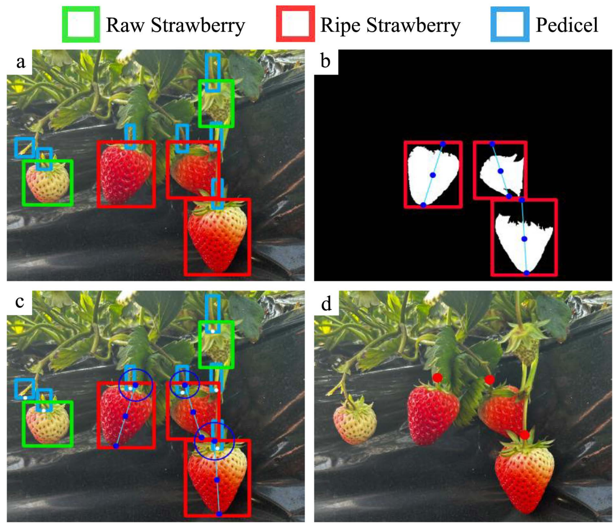

Fig.9 Harvest target prediction results

Notes: (a) recognition result; (b) apex of the fruit; (c) identification of the affiliated fruit pedicel; (d) harvest target position result.

2.3 Comparison with other detection algorithms

In this study, an improved YOLOv8 object detection model was proposed to accurately and swiftly recognize strawberry fruits and pedicels in various complex scenarios, and position the picking points of ripe strawberry fruits.

The results of this study were compared with other studies, as shown in Table 5. Traditional machine vision methods for strawberry fruit recognition and picking point localization are suitable only for simpler scenarios and lack applicability to complex picking situations[29,30]. Deep learning methods like Mask-RCNN and improved MRS Faster R-CNN have been proposed for recognizing strawberry fruits and determining picking points. However, these methods fell short in both recognition speed and accuracy[12,18]. Another approach introduced the ASFF module in YOLOv5 to enhance strawberry image feature extraction, but did not outperform the method proposed in this study[15]. Overall, the method presented in this study achieved the highest recognition accuracy and shortest processing time among the compared methods.

Table 5 Comparison of YOLO8-SCE and other methods of strawberry detection

Method	Fruit features	Picking oint	Recognition erformance	Time
Machine vision (color)	Occlusion,separation	√	Accuracy>76.55%	—
Machine vision (color and shape)	separation	√	Accuracy>98.30%	0.89 s
MRS Faster R-CNN	Occlusion,separation,overlapping	-	F1:95.29	—
Mask R-CNN	Occlusion,separation,overlapping	√	F1:89.85	0.13 s
YOLOv5-ASFF	Occlusion,separation,overlapping	-	F1:88.03	0.02 s
Our Method	Occlusion,separation,overlapping	√	F1:98.62	0.10 s

2.4 Practical Deployment and Performance Verification

To validate whether the proposed YOLOv8-SCE model meets the requirements for efficient and accurate target recognition in the visual system of a strawberry harvesting robot, the model was deployed on a strawberry harvesting robot equipped with a robotic arm, as shown in Fig.10. The robot is equipped with an RGB-D camera (RealSense D435i) capable of capturing color images of ground-planted strawberries. The robot synchronously moves with the robotic arm and the harvesting end-effector during operation.

Fig.10 The strawberry picking robot

Table 6 Experimental results of YOLOv8-SCE model deployed on the strawberry picking robot

Target	Detect	Correct	Accuracy/%
Strawberry Fruit	118	116	98.30
Strawberry Pedicel	164	151	92.07
Harvest Point Localization	72	68	94.44

To evaluate the recognition performance of the improved model in real-world scenarios and the harvesting performance of the integrated robot system, tests were conducted in strawberry-planted ridge environments. The test scenarios included various occlusion types. A total of 118 strawberries (including 72 mature fruits) were tested, and the causes of harvesting failures were statistically analyzed. The primary failure reasons included missed detections, false detections (e.g., misclassifying immature strawberries as mature or incorrect category identification), and failures in harvest point localization. Table 6 presents the statistical results, demonstrating that the deployed DSW-YOLO model successfully detected most strawberries while achieving a relatively high success rate in harvest point localization.

In conclusion, the YOLOv8-SCE model exhibits excellent generalization capability and robustness, enabling accurate strawberry detection and harvest point localization, thereby fulfilling the precision requirements of the visual system for strawberry harvesting robots.

3 Conclusions

A lightweight YOLOv8-SCE model and a strawberry picking point localization method was proposed. The model achieves a detection speed of 38.44 FPS, with mAP@0.5 and precision rates of 99.04% and 99.23% respectively. It demonstrates stable performance in both fruit occlusion and complex background scenarios, showing significant pedicel recognition accuracy. The proposed picking point localization method, based on recognition results, achieves 91.11% localization accuracy in independent tests, with a single fruit processing time of 96 ms. Physical robot tests confirm that the recognition and localization accuracy exceeds 92.07%, meeting practical harvesting requirements. The method is suitable for non-severe pedicel overlap scenarios. Future work will focus on enhancing dataset complexity and optimizing algorithms to improve anti-interference capabilities.

References

[1] WARNER R, WU B S, MacPherson S, et al. A review of strawberry photobiology and fruit flavonoids in controlled environments [J]. Frontiers in Plant Science, 2021, 12: 611893.

[2] HERNÁNDEZ-MARTÍNEZ N R, BLANCHARD C, WELLS D, et al. Current state and future perspectives of commercial strawberry production: A review [J]. Scientia Horticulturae, 2023, 312: 111893.

[3] ALI M M, ANWAR R, MALIK AU, et al. Plant growth and fruit quality response of strawberry is improved after exogenous application of 24-epibrassinolide [J]. Journal of Plant Growth Regulation, 2023, 41: 1786-1799.

[4] DEFTERLI S G. Review of robotic technology for strawberry production [J]. Applied Engineering in Agriculture, 2016, 32: 301-18.

[5] YU Y, ZHANG K, LIU H, et al. Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot [J]. IEEE Access, 2020, 8: 116556-116568.

[6] XIONG Y, GE Y, GRIMSTAD L, et al. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation [J]. Journal of Field Robotics, 2020, 37: 202-224.

[7] LI Z, YUAN X, WANG C. A review on structural development and recognition-localization methods for end-effector of fruit-vegetable picking robots [J]. International Journal of Advanced Robotic Systems, 2020, 19: 17298806221104906.

[8] TANG Y, CHEN M, WANG C, et al. Recognition and localization methods for vision-based fruit picking robots: A review [J]. Frontiers in Plant Science, 2020, 11: 510.

[9] ZHANG J, KANG N, QU Q, et al. Automatic fruit picking technology: a comprehensive review of research advances [J]. Artificial Intelligence Review, 2024, 57: 54.

[10] HABARAGAMUWA H, OGAWA Y, SUZUKI T, et al. Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network [J]. Engineering in Agriculture, Environment and Food, 2018, 11: 127-138.

[11] FUJINAGA T. Strawberries recognition and cutting point detection for fruit harvesting and truss pruning [J]. Precision Agriculture, 2024, 25: 1262-1283.

[12] ZHANG Y, ZHANG L, YU H, et al. Research on the strawberry recognition algorithm based on deep learning [J]. Applied Sciences, 2023, 13: 11298.

[13] ZHANG Y, YU J, CHEN Y, et al. Real-time strawberry detection using deep neural networks on embedded system (rtsd-net): An edge AI application [J]. Computers and Electronics in Agriculture, 2022, 192: 106586.

[14] TAO Z, LI K, RAO Y, et al. Strawberry maturity recognition based on improved YOLOv5 [J]. Agronomy, 2024, 14(3): 460.

[15] LI Y, XUE J, ZHANG M, et al. YOLOv5-ASFF: A multistage strawberry detection algorithm based on improved YOLOv5 [J]. Agronomy, 2023, 13: 1901.

[16] CHEN Y, XU H, CHANG P, et al. CES-YOLOv8: Strawberry maturity detection based on the improved YOLOv8 [J]. Agronomy, 2024, 14: 1353.

[17] WANG C, HAN Q, LI C, et al. Assisting the planning of harvesting plans for large strawberry fields through image-processing method based on deep learning. agriculture [J]. Agriculture, 2024, 14: 560.

[18] YU Y, ZHANG K, YANG L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN [J]. Computers and Electronics in Agriculture, 2019, 163: 104846.

[19] WANG F, URQUIZO R C, ROBERTS P, et al. Biologically inspired robotic perception-action for soft fruit harvesting in vertical growing environments [J]. Precision Agriculture, 2023, 24: 1072-1096.

[20] RONG Q, HU C, HU X, et al. Picking point recognition for ripe tomatoes using semantic segmentation and morphological processing [J]. Computers and Electronics in Agriculture, 2023, 210: 107923.

[21] XIAO B, NGUYEN M, YAN W Q. Fruit ripeness identification using YOLOv8 model. multimedia tools and applications [J]. Multimedia Tools and Application, 2024, 83: 28039-28056.

[22] YANG S, WANG W, GAO S, et al. Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-swin transformer [J]. Computers and Electronics in Agriculture, 2023, 215: 108360.

[23] WANG Q, YANG L, ZHOU B, et al. YOLO-SS-Large: a lightweight and high-performance model for defect detection in substations [J]. Sensors, 2023, 23: 8080.

[24] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: A survey. Computational visual media [J]. Computational, 2022, 8: 331-368.

[25] NIU Z, ZHONG G, YU H. A review on the attention mechanism of deep learning [J]. Neurocomputing, 2021, 452: 48-62.

[26] CHEN J, MA A, HUANG L, et al. Efficient and lightweight grape and picking point synchronous detection model based on key point detection [J]. Computers and Electronics in Agriculture, 2024, 217: 108612.

[27] ZOU H, WANG Z. An enhanced object detection network for ship target detection in SAR images [J]. The Journal of Supercomputing, 2024, 80: 17377-17399.

[28] LI J, LI J, ZHAO X, et al. Lightweight detection networks for tea bud on complex agricultural environment via improved YOLO v4 [J]. Computers and Electronics in Agriculture, 2023, 211: 107955.

[29] RAJENDRA P, KONDO N, NINOMIYA K, et al. Machine vision algorithm for robots to harvest strawberries in tabletop culture greenhouses [J]. Engineering in Agriculture, Environment and Food, 2009, 2: 24-30.

[30] WANG L, ZHANG L, DUAN Y, et al. Fruit localization for strawberry harvesting robot based on visual servoing [J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31: 25-31.