# YOLOv8-STEM: ENHANCED OVERHEAD APPLE STEM DETECTION UNDER OCCLUSIONS

/

# YOLOv8-STEM：俯视视角下的苹果果柄遮挡识别

**Li WANG[1)], Yanqi SUN[1)], Tianle ZHANG[1)], Panpan YAN[2)], Qiangqiang YAO[3)], Zhen MA[4)], Degui MA[5)], Xingdong SUN[1*)]**

[1)]Anhui Agricultural University College of Engineering, Hefei, Anhui, China, 230036
[2)] Heze Vocational College, Heze,Shandong, China, 274002
[3)]Qinghai University, China, 274002, 810016
[4)]School of Agricultural Engineering, Jiangsu University, Zhenjiang, Jiangsu, China, 212000
[5]Anhui Agricultural University College of Engineering，Hefei, Anhui, China, 230036

*Corresponding author: Xingdong Sun*
*Tel: 0551-65786450; E-mail:xdsun@ahau.edu.cn*
*DOI: https://doi.org/10.35633/inmateh-76-07*

***Keywords:*** *stem detection, improved yolov8, occlusion recognition, orchard*

## ABSTRACT

*Accurate detection of apple stems is crucial for robotic cutting. This study proposed an improved YOLOv8-stem method for apple stem detection in overhead imagery under occlusion conditions. First, several improvements were made to the YOLOv8 neural network: the conventional convolutional process within the intermediate neck layer was substituted with the AK Convolution mechanism, a small object detection head was added, and ResBlock+CBAM attention mechanism was incorporated. Second, stem occlusion was determined by analyzing the positional relationship between the detected bounding boxes of stems and apples. The experimental results showed that compared to the original YOLOv8, this method improved apple stem detection accuracy by 6.0% (from 79.9% to 85.9%) and increased harvesting completeness from 84.2% to 93.2%.*

## 摘要

*准确检测苹果茎对于机器人切割至关重要。本研究提出了一种改进的 YOLOv8 茎检测方法，用于遮挡条件下俯视图像中的苹果茎检测。首先，对 YOLOv8 神经网络进行了几项改进：用 AK 卷积机制取代了中间颈层内的传统卷积过程，增加了一个小目标检测头，并引入了 ResBlock+CBAM 注意力机制。其次，通过分析检测到的茎和苹果边界框之间的位置关系来确定茎的遮挡。实验结果表明，与原始 YOLOv8 相比，该方法将苹果茎检测准确率提高了 6.0%（从 79.9%提高到 85.9%），收获完整性从 84.2%提高到 93.2%。*

## INTRODUCTION

Apples are the third largest fruit in terms of area planted and production globally *(Vasylieva et al., 2021)*. Manual apple harvesting is a time-consuming and labor-intensive task, and improving the competitiveness of the apple market requires addressing this challenge. Leaving apple stem too long during harvesting can cause surface scratches on the apples during transportation and storage, which can affect their freshness and appearance. Fruits stored without their stems exhibited greater weight loss, higher decay rates, and increased vitamin C loss *(Ozturk et al., 2020)*. Small apple stems and complex growing environments, overlapping fruits, branch and leaf shading, and light variations pose challenges to accurate target identification *(Chen et al., 2019)*.

In recent years, a lot of related works had been done by scholars at home and abroad for small target recognition of fruit stems in complex orchard environments. The YOLO family of algorithms had become a leader in target detection due to its rapid development and excellent performance. However, YOLO algorithms were mainly designed to detect and recognize full-size objects, and their performance was not as good when facing special-size objects, especially small targets *(Liu et al., 2023; Liu et al., 2022)*. To address this issue, Wu et al. introduced the YOLO-Banana model for precise identification of bananas and to determine the fruit axis and cutting point. They enhanced the Bottleneck module of YOLOv5 and employed an edge-detection algorithm to segment the fruit axis's contour, allowing them to pinpoint the cut-off location *(Wu et al., 2022; Wu et al., 2021)*. In situations where the fruit stem was covered, *Yu et al.* enhanced Mask R-CNN to address the issues of limited robustness in traditional algorithms when operating in unstructured environments *(Yu et al., 2019)*.

To investigate the method for identifying picking points in cases of partial occlusion, *Xiong et al. (2018)* conducted a study on locating grapes in the presence of disturbances. Other researchers employed stereovision and image processing techniques to detect and determine the locations of grape clusters and picking points, subsequently performing size measurements and enclosure calculations for the grapes *(Luo et al., 2021; Lufeng et al., 2017)*. The analysis of these studies revealed that there were fewer network improvements for short targets such as apple fruit stems, and for the localization of picking points covering the stems, most of them directly segmented the image of the target to be picked, which could produce large errors in the localization of the picking points.

This study concentrated on the detection of apple stems and sought to enhance the conventional YOLOv8 algorithm *(Jocher et al., 2023)*. It presented an enhanced YOLOv8-based method designed specifically for detecting apple stems in overhead imagery. To achieve this, high-quality datasets of apple fruits and stems were collected from a top-down perspective to provide positive samples for network training. In the annotation of training targets, both apples occluded by tree branches and apple stems were annotated as separate categories. When utilizing YOLOv8 for the detection of apples occluded by branches, the presence of an apple stem within the annotated apple bounding box was verified. If a stem was present, it was recognized normally; if not, it was determined that the stem was occluded by branches, necessitating a modification of the camera's position and angle. The YOLOv8 architecture was enhanced by incorporating a detection head specifically designed for small objects, the conventional convolutional process within the intermediate neck layer was substituted with the AK Convolution *(Zhang et al., 2023)* mechanism introducing a 160×160 detection feature map. A ResBlock+CBAM attention mechanism *(Woo et al., 2018)* module was incorporated to enhance the model's feature extraction capabilities for small target objects, thereby adapting to the characteristics of the apple stem dataset. The mean average precision (mAP) value and the harvesting completeness were calculated to evaluate the effectiveness of the stem detection, where the harvesting completeness was determined by applying the detection algorithm to apple stem images and computing the ratio of successfully detected stems to the total number of stems present in the dataset. This study provided a reference for designing a precise cutting vision system for harvesting robots.

## MATERIALS AND METHODS

### *Collection and preprocessing of apple stem data*

This study investigated the real orchard environment and the growth of apple stems in Shandong, China, and collected an apple dataset.

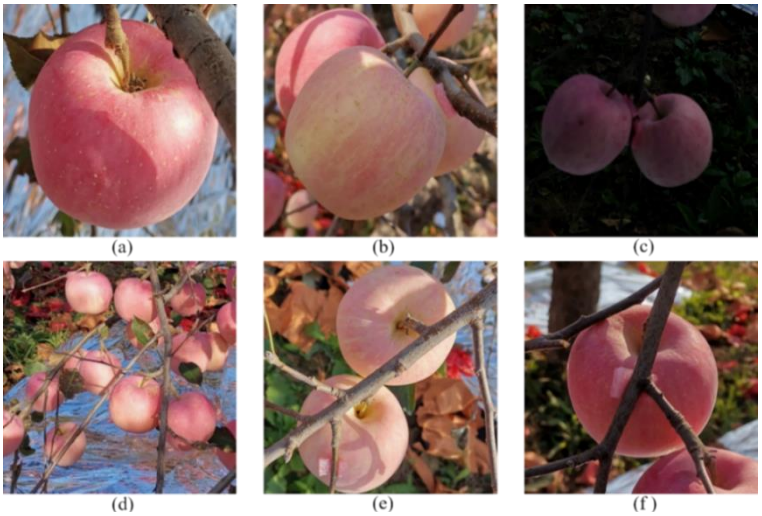


**Fig. 1 - Apple images in different conditions**

The images of apple stems were collected from apple orchards in Yantai, Shandong, China on October 15th and 16th, 2023. In order to minimize the risk of overfitting the network model, this study collected different distances, including close range (0.3 meters to 1 meter) and long range (1 meter to 3 meters). Images were taken at different times (morning, noon, and evening) to obtain rich and diverse data. Finally, the image dataset was cropped and organized. It was concluded that the apple stem's integrity was highest when viewed from an overhead (top-down) or horizontal perspective. A total of 1671 images of apple stems were collected.

As shown in Figure 1, this was a typical set of apple stem images in a complex environment, this study also collected images of fruit stems completely covered to cope with the special situation of complex orchard environments. All the experimental images were adjusted to a uniform size of 640×640, and then the image transformation module provided by the PyTorch framework was used to enhance the data.

***Improved YOLOv8-stem Network Architecture***

In order to accurately and quickly identify small objects such as apple stems, this study replaced the original convolutions with AKConv in the 3th and 5th layers of the network, based on the YOLOv8n model. Additionally, it incorporated a 160×160 small object detection head to increase sensitivity to smaller targets. Integrating the CBAM attention mechanism into the P4 and P5 feature maps improved feature representation, suppressed irrelevant information, improved object detection performance, and boosted the network's adaptability to objects of varying scales. The following sections will detail the working principles and technical aspects of each module. The altered network architecture is shown in Figure 2.

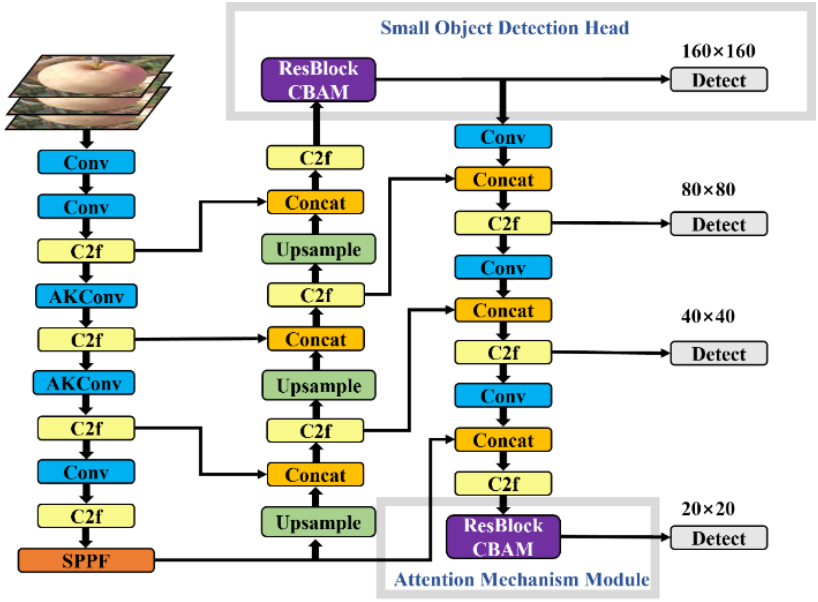


**Fig. 2 - YOLOv8-stem Network**

***Replace the convolution module with AKConv***

AKConv assigned initial sampling coordinates to convolutions of varying sizes and modifies the sample shape using learnable offset values. The third and fifth layers typically represent the deep feature extraction stage of the network, with higher resolution and channel numbers. Replacing AKConv in these layers can help the network focus more on extracting high-level and abstract features, thereby enhancing the network's capability to extract features from apple stems and manage occlusion scenarios.

***Adding Small Target Detection Head***

In the standard YOLOv8 object detection model, the output consists of three layers: P3, P4, and P5, each with three default detection heads. However, the standard model may perform poorly in detecting small objects. Therefore, a new 160x160 detection feature map was introduced for detecting objects larger than 4x4.

***ResBlock＋CBAM***

When using CBAM in ResNet improved the expressiveness of the feature maps in both the channel and spatial dimensions, thus enhancing the performance of the model. The high-level semantic information and multi-scale information of P4 and P5 made them particularly suitable for detecting small objects. By incorporating the CBAM attention mechanism into the P4 and P5 layers, the model's ability to represent features was enhanced, irrelevant information was suppressed, and target detection performance was improved, thus improving the model's detection and recognition capabilities for apple stems. This weighting mechanism allows the model to be more generalizable, adapting to apple stems of different sizes, shapes, and positions, thereby improving the model's robustness and accuracy.

***Experimental Results and Analysis***

In experiments to improve various backbone networks, DSConv *(Gennari et al., 2019)* convolution provided better accuracy and continuity in the segmentation of medical tubular structures, but the accuracy of apple stem detection decreased by 31.3%. Replacing the backbone network with ODConv *(Li et al., 2022)* moderately improved accuracy by 0.3% without changing the number of parameters. The results indicated that neither DSConv nor ODConv could effectively extract fruit stem features in highly crowded and complex environments. Bifpn was used to achieve bidirectional fusion of deep and shallow features, but the model parameters and accuracy of identifying fruit stems remained almost unchanged *(Tan et al., 2020)*. The CARAFE *(Wang et al., 2019)* network, which incorporated a kernel prediction module and a content-aware recombination module to generate larger up-sampling kernels, showed a decrease in accuracy of 0.25%. Vanillanet *(Chen et al., 2024)* performed well in terms of parameter and model computation by scaling the neural network model, but its accuracy control was mediocre. On the other hand, the improvement by replacing the backbone network with AKConv showed excellent performance, with a slight decrease in parameters and model computation while increasing detection accuracy (mAP@0.5/% increased by 2.2). This method effectively balanced model parameters and performance. Experimental results are shown in Table 1.

**Table 1**

**Detection accuracy of each model in dataset (%)**

| Backbone | mAP@0.5/% | mAP@0.5-0.95/% | Params/M | FLOPs/G |
|---|---|---|---|---|
| Yolov8n(base) | 0.799 | 0.362 | 3011027 | 8.2 |
| Yolov8- DSConv | 0.486 | 0.181 | 2996959 | 29.6 |
| Yolov8-ODConv | 0.802 | 0.359 | 2999036 | 8.1 |
| Yolov8- Bifpn | 0.804 | 0.342 | 3005852 | 8.1 |
| Yolov8- CARAFE | 0.774 | 0.332 | 4053191 | 8.3 |
| Yolov8-Vanillanet | 0.789 | 0.334 | 1731635 | 5.0 |
| Yolov8- AKConv | **0.821** | **0.363** | **2965337** | **8.0** |

Among the improvements in various attention mechanisms, LSKA *(Lau et al., 2024)* did not improve accuracy but reduced the model's parameter count. The EMA *(Ouyang et al., 2023)* module focused on retaining feature information on each channel, resulting in a slight increase in accuracy by 2.1%, with no increase in model computation. The parameter-free attention module SE *(Hu et al., 2018)* showed a small improvement in accuracy, while the global attention mechanism GAM *(Liu et al., 2021)* demonstrated a modest rise in the number of parameters while achieving an accuracy improvement of 3.9%. CBAM widely applied in many improvement studies, effectively improved accuracy in this study by combining feature channel and spatial principles to adapt to the complex and multi-feature environment in this scenario. Compared to the base accuracy, mAP@0.5/% improved by 4.6%, and the parameter count and the model's computations were roughly equivalent to those of the original model. Experimental results are shown in Table 2.

**Table 2**

**Various attention mechanisms**

| Attention mechanism | mAP@0.5/% | mAP@0.5-0.95/% | Params/M | FLOPs/G |
|---|---|---|---|---|
| Yolov8n(base) | 0.799 | 0.362 | 3011027 | 8.2 |
| Yolov8-C2f-LSKA | 0.794 | 0.338 | 2351555 | 6.6 |
| Yolov8n+EMA | 0.825 | 0.355 | 3006515 | 8.1 |
| Yolov8n+SEAttention | 0.808 | 0.368 | 3227923 | 8.4 |
| Yolov8n+GAM | 0.838 | 0.337 | 3687123 | 9.5 |
| **Yolov8n+CBAM** | **0.845** | **0.346** | **3014406** | **8.1** |

After testing a variety of attentional mechanisms for detecting small objects, the parts with significant improvements were selected and incorporated. Comparative experiments revealed that the YOLO network algorithm demonstrated strengths with respect to the three improvement modules from current cutting-edge research. However, additional ablation experiments were required to confirm the efficacy of each enhancement module. The findings from these experiments could be found in Table 3.

Table 3

Module ablation experiment

| AKConv | CBAM | Head | mAP@0.5/% | mAP@0.5-0.95/% | Params/M | FLOPs/G |
|---|---|---|---|---|---|---|
| ✓ | | | 0.821 | 0.362 | 2965337 | 8.0 |
| | ✓ | | 0.845 | 0.342 | 3014406 | 8.1 |
| | | ✓ | 0.826 | 0.352 | 2977588 | 12.5 |
| | ✓ | ✓ | 0.835 | 0.347 | 3239726 | 8.9 |
| ✓ | ✓ | ✓ | 0.859 | 0.347 | 3998948 | 14.6 |

These three improvements enabled the YOLO model to achieve higher recognition accuracy for small objects with specific apple stem shapes in orchard environments, where occlusion and scene complexity are common. Notably, this enhancement in performance was achieved with only a minimal increase in the number of parameters. Ultimately, compared with the baseline model, the improved YOLO achieved a 6.0% increase in mAP@0.5 on the dataset. The results of the ablation experiments are presented in Table 5. A comparison of detection outcomes before and after the network modifications, as shown in Figure 3, illustrates the effectiveness of the architectural changes. The visual results highlight clear improvements in detection accuracy and robustness achieved through the proposed enhancements.

### *Evaluation metrics*

This study comprehensively evaluates the model's performance using precision ($P$), recall ($R$), gigafloating-point operations per second (GFLOPs), and the mean Average Precision (mAP) across all target categories at an IoU threshold of 0.5. The formulas for calculating precision, recall, and average precision (AP) are as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 P(R)dr \tag{3}$$

The mean Average Precision (mAP) is calculated as the total of the average precisions for all labels divided by the number of classes.

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k(classes)} \tag{4}$$

Where $n$ denotes the total number of categories, $k$ is the number of detections, and $AP$ represents the average precision for each category.

## RESULTS

### *Experimental Environment*

Experiment setup: The operating system of the experimental platform was Windows 10. The CPU was an Intel(R) Core (TM) i5-12490F, with 16GB of RAM. The GPU was an Nvidia GeForce RTX 4060 Ti with 8GB of VRAM. The CUDA version used by torch vision is 11.7, and the deep learning framework was PyTorch-GPU 2.0.1+cu118. To ensure the fairness and rationality of the experiments, all ablation experiments were conducted in the same experimental environment.

### *Dataset Preparation*

In apple stem images captured from various overhead angles, single-angle camera detection may fail to accurately identify the stem when the field of view is partially obstructed by overlapping branches or fruits. Such occlusions can prevent the camera from capturing the full shape and position of the apple stem, negatively impacting subsequent harvesting or processing tasks. To address this challenge, a specialized labeling strategy was adopted for the dataset used in this study. Apple stems visible from a top-down perspective were labeled as *Apple stem*, while apples obscured by branches were labeled as *Masked apple*. Additionally, apple stems captured from the top-down view were annotated together with apple eyelets to enhance the feature representation of the stem. This custom-labeled dataset formed the basis for training and evaluating the proposed detection algorithm designed to identify apple stems in occluded environments.

This study trained a model to recognize two types of objects: apple stems and apples obscured by tree branches. After detection, it was further determined whether the bounding box of the obscured apple contains the coordinates of the apple stem. If the coordinates of the stem are not within the bounding box of the obscured apple, it could be concluded that the stem is obscured by a branch. Otherwise, it was considered exposed. When tree branches obstruct the apple and cover the stem, the detection angle and position could be adjusted to better observe the stem, thereby more accurately detecting the initially unrecognized stem and providing a technical solution for apple harvesting.

There were a total of 2816 *Apple stem* labeled targets and 1321 *Masked apple* labeled targets in the original 1671 images before data augmentation. It should be noted that because the *Masked apple* was larger and more distinctive in the images, it was not involved in the statistics of the YOLOv8 network structure optimization results. The performance evaluation was conducted only for the target recognition effect of the *Apple stem*. The training of the *Masked apple* category was performed to validate and provide statistics for the specific case when the apple was covered by branches and the fruit stem was also covered. The dataset categorization is shown in Table 4.

**Table 4**

**Classification of data sets**

| Classification of data sets | Number | Use of data sets |
|---|---|---|
| Apple stem | 2816 | Network Modification Performance Evaluation |
| Masked apple | 1321 | Identify and count obscured apples |

This study employed K-Fold cross-validation as the evaluation method for the machine learning model. The original dataset was divided into 5 subsets (K=5). In each iteration, one subset was used as the validation set, while the remaining four subsets (K-1) were used as the training set. This process was repeated five times, ensuring that each subset served as the validation set exactly once. After each iteration, the model's performance metrics were calculated on the validation set, and the average of these results was taken as the overall performance evaluation on the entire dataset. The training results presented in the subsequent experiments were obtained using this K-Fold cross-validation approach.

***Validate the network model***

In order to more intuitively test the apple stem detection performance of the modified network model, this study compared the apple stem detection performance of YOLOv8n and YOLOv8 stalk. The comparison of the detection results is shown in Figure 3.


**Fig. 3 - The comparison of detection results**

There are a total of three sets of apple stem detection comparisons in the comparison chart. Overall, in terms of detection performance, the improved YOLOv8 stalk has increased the recall rate and overall confidence of apple stem detection.

After comparing the original YOLOv8 model, this study validated the superiority of the improved YOLOv8 stalk model by training and testing it on the same dataset as other detection models such as faster R-CNN, SSD, YOLOv5, YOLOv3, etc. From Table 5, it can be seen that the improved YOLOv8 stalk model has the highest average accuracy compared to other models, with an increase of 5.4 percentage points in average accuracy compared to the original YOLOv8n network.

**Table 5**

**Comparison of the effectiveness of different network models**

| model | mAP @0.5/% | mAP @0.5-0.95/% | Params/M | FLOPs/G |
|---|---|---|---|---|
| YOLOv3-tiny | 76.2 | 33.6 | 4.3 | 18.9 |
| YOLOv5n | 80.1 | 35.7 | 1.9 | 4.5 |
| YOLOv8n | 80.3 | 36.0 | 3.0 | 8.2 |
| Faster R-CNN | 80.8 | 35.8 | 104 | 268.3 |
| SSD | 82.2 | 35.4 | 93 | 195.3 |
| YOLOv8-stalk | 85.7 | 36.7 | 4.0 | 14.6 |

***Detection Results and Analysis of Occluded Stems***

After modifying the network, the deep learning model achieved higher recognition accuracy for apple stems, but it could not address the issue of incomplete harvesting caused by branches obscuring the apple stems. Therefore, the modified network was used to detect cases where apple stems were obscured by branches to observe the recognition and improvement effects. This study categorized the cases of obscured apple stems into three types: stems inside the corresponding obscured apple box but not recognized, stems inside the corresponding obscured apple box and recognized, and stems inside a non-corresponding apple box. This last category included cases where the stem was recognized or not recognized within the non-corresponding apple box. The number of stems not recognized was the focus for angle adjustment and re-detection to improve the actual harvesting rate. The following images show the detection results for these three different scenarios. Different occlusion scenarios of the stems are shown in Figure 4.


**Fig. 4 - Different occlusion scenarios of the stems**

This experiment was conducted using the trained inference model on 428 new scene images of apple stems collected from an apple orchard, which were different from the training dataset. Among these new images, 576 apples were annotated as obscured apples (Actual number of occluded apples). After prediction by the model, there were 565 predicted obscured apples (Predicted number of occluded apples). The number of stems recognized within the corresponding obscured apple bounding boxes (Fruit stems not covered by branches) was 485, and the number of stems inside the corresponding obscured apple boxes but not recognized (Fruit stems obstructed by branches) was 52. Other cases included 28 stems inside non-corresponding apple boxes (Other situations).

The statistical results are presented in Table 6.

**Table 6**

**The statistical results**

| Category | number |
|---|---|
| Images of apple orchard scenes | 428 |
| Actual number of occluded apples | 576 |
| Predicted number of occluded apples | 565 |
| Fruit stems not covered by branches | 485 |
| Fruit stems obstructed by branches | 52 |
| Other situations | 28 |

After analyzing the data of predicted obscured apples in the apple orchard scene images, in the case of 576 actually obscured apples, if the stems are also obscured, the deep learning network will not detect the obscured apples from a single camera angle, resulting in 52 stems inside the corresponding obscured apple box but not recognized. When this situation is detected, it can be inferred that the stems are obscured. In this case, changing the position and angle of the single camera to re-recognize the obscured stems can improve the picking efficiency from 576/485=84.2% to 576/ (485+52) =93.2% according to the theoretical statistical data. Moreover, as the apples with recognized stems inside the corresponding obscured apple box are picked, the situation where the stems are recognized inside non-corresponding obscured apple boxes will decrease, further improving the picking efficiency.

## CONCLUSIONS

This study presented an enhanced YOLOv8-stem approach for detecting apple stems in overhead imagery under occlusion conditions. First, two datasets comprising apples and apple stems obstructed by tree branches were collected. Second, the YOLOv8 neural network was improved by substituting the conventional convolutional process within the intermediate neck layer with the AK Convolution mechanism. A small object detection head was added, introducing a 160×160 detection feature map, and a ResBlock+CBAM attention mechanism was incorporated. Third, the likelihood of missed detections was decreased by utilizing the coordinates of the apples and apple stems obstructed by tree branches identified by the enhanced YOLOv8-stem model. The experimental results showed that compared to the original YOLOv8 model, this method significantly improved both the detection accuracy of apple stems and the completeness of harvesting. The enhanced stem detection model achieved an average precision of 85.9%, representing a 6.0% improvement over the baseline YOLOv8 model (from 79.9% to 85.9%), and the harvesting completeness increased from 84.2% to 93.2%. This approach demonstrated promising applications in precise apple harvesting and automated fruit picking systems.

## ACKNOWLEDGEMENT

This research was funded by the National Natural Science Foundation of China (No.52375228); the National Natural Science Foundation of China (No.52005009); the Post-Master's Enterprise Workstation in Anhui Province (No.2022sshqygzz012); the Key Projects of Natural Science Research Projects of Colleges and Universities in Anhui Province, China (No.KJ2021A0154); High-tech Key Laboratory of Agricultura Equipment and Intelligence of Jiangsu Province (MAET202318), College of Agricultural Engineering, Jiangsu University (MAET202318).

## REFERENCES

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

[2] Chen Y., Lee W. S., Gan H., Peres N., Fraisse C., Zhang Y., & He, Y. (2019). Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages (基于深度神经网络的草莓产量预测), *Remote Sensing*, 11(13), 1584. DOI:10.3390/rs11131584

[3] Chen H., Wang Y., Guo J., & Tao, D. (2024). Vanillanet: the power of minimalism in deep learning (Vanillanet：深度学习中极简主义的力量), *Advances in Neural Information Processing Systems*, 36.

[4] Gennari M., Fawcett R., & Prisacariu V.A. (2019). DSConv: Efficient convolution operator, *arxiv preprint* arxiv:1901.01928. DOI: 10.48550/arXiv.1901.01928

[5] Hu J., Shen L., & Sun G. (2018). Squeeze-and-excitation networks (挤压和激励网络), In *Proceedings of the IEEE conference on computer vision and pattern recognition.* pp.7132-7141. DOI: 10.48550/arXiv.1709.01507

[6] Jocher G., Chaurasia A., & Qiu J. (2023). *Ultralytics YOLOv8*. Retrieved from https://github.com/ultralytics/ultralyticsw

[7] Lau K.W., Po L.M., & Rehman Y.A.U. (2024). Large separable kernel attention: Rethinking the large kernel attention design in CNN, *Expert Systems with Applications*, 236, 121352. DOI:10.1016/j.eswa.2023.121352

[8] Li C., Zhou A., & Yao A. (2022). Omni-dimensional dynamic convolution (全维动态卷积), *arxiv preprint* arxiv:2209.07947. DOI:10.48550/arXiv.2209.07947

[9] Liu H., Duan X., Chen H., Lou H., & Deng L. (2023). DBF-YOLO: UAV Small Targets Detection Based on Shallow Feature Fusion (DBF-YOLO：基于浅层特征融合的无人机小目标检测), *IEEJ Transactions on Electrical and Electronic Engineering*, 18(4), 605-612. DOI:10.1002/tee.23758

[10] Liu H., Sun F., Gu J., & Deng L. (2022). Sf-YOLOv5: A lightweight small object detection algorithm based on improved feature fusion mode (Sf-YOLOv5：一种基于改进特征融合模式的轻量级小目标检测算法), *Sensors*, 22(15), 5817. DOI:10.3390/s22155817

[11] Liu Y., Shao Z., & Hoffmann N. (2021). Global attention mechanism: Retain information to enhance channel-spatial interactions (全球注意力机制：保留信息以增强渠道空间互动), *arxiv preprint* arxiv:2112.05561. DOI:10.48550/arXiv.2112.05561

[12] Lu feng L., Xiang jun Z., Cheng lin W., Xiong C., Zi Shang Y., & Weiming S. (2017). Recognition method for two overlapping and adjacent grape clusters based on image contour analysis (基于图像轮廓分析的重叠相邻葡萄簇识别方法), *Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery*. 48(6).

[13] Luo L., Liu W., Lu Q., Wang J., Wen W., Yan D., & Tang Y. (2021). Grape berry detection and size measurement based on edge image processing and geometric morphology (基于边缘图像处理和几何形态学的葡萄果实检测与尺寸测量), *Machines*, 9(10), 233. DOI:10.3390/machines9100233

[14] Ouyang D., He S., Zhang G., Luo M., Guo H., Zhan J., & Huang Z. (2023), June. Efficient multi-scale attention module with cross-spatial learning (具有跨空间学习的高效多尺度注意力模块), In ICASSP 2023-2023 IEEE *International Conference on Acoustics, Speech and Signal Processing* (ICASSP). pp.1-5. IEEE. mDOI:10.1109/ICASSP49357.2023.10096516

[15] Ozturk B., Aglar E., Gun S., & Karakaya O. (2020). Change of fruit quality properties of jujube fruit (Ziziphus jujuba) without stem and with stem during cold storage, *International Journal of Fruit Science*, 20(sup3), S1891-S1903. DOI: 10.1080/15538362.2020.1834901

[16] Tan M., Pang R., & Le Q. V. (2020). Efficientdet: Scalable and efficient object detection (Efficientdet：可扩展且高效的对象检测), *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781-10790). DOI:10.48550/arXiv.1911.09070

[17] Vasylieva N., & Harvey J. (2021). Production and trade patterns in the world apple market, *Innovative Marketing*, 17(1), 16. DOI: 10.21511/im.17(1).2021.02

[18] Wang J., Chen K., Xu R., Liu Z., Loy C. C., & Lin D. (2019). Carafe: Content-aware reassembly of features (Carafe: 功能的内容感知重组), *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3007-3016).

[19] Woo S., Park J., Lee J. Y., & Kweon I. S. (2018). Cbam: Convolutional block attention module, *In Proceedings of the European conference on computer vision (ECCV)* (pp. 3-19). DOI: 10.48550/arXiv.1807.06521

[20] Wu F., Duan J., Ai P., Chen Z., Yang Z., & Zou X. (2022). Rachis detection and three-dimensional localization of cut off point for vision-based banana robot (基于视觉的香蕉机器人种族检测和切割点三维定位), *Computers and Electronics in Agriculture*, 198, 107079. DOI: 10.1016/j.compag.2022.107079

[21] Wu F., Duan, J., Chen S., Ye Y., Ai P., & Yang Z. (2021). Multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point (香蕉多目标识别及花序轴切割点自动定位), *Frontiers in plant science*, 12, 705021. DOI:10.3389/fpls.2021.705021

[22] Xiong J., He Z., Lin R., Liu Z., Bu R., Yang Z., Peng H., & Zou X. (2018). Visual positioning technology of picking robots for dynamic litchi clusters with disturbance (动态荔枝丛采摘机器人的视觉定位技术), *Computers and electronics in agriculture*. 151, 226-237. DOI: 10.1016/j.compag.2018.06.007

[23] Yu Y., Zhang K., Yang L., & Zhang D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN (基于 Mask RCNN 的非结构化草莓采摘机器人果实检测), *Computers and electronics in agriculture*, 163, 104846. DOI: 10.1016/j.compag.2019.06.001

[24] Zhang X., Song Y., Song T., Yang D., Ye Y., Zhou J., & Zhang L. (2023). AKConv: Convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters (AKConv：具有任意采样形状和任意数量参数的卷积核), *arxiv preprint arxiv:2311.11587*. DOI: 10.48550/arXiv.2311.11587