YOUNG APPLE FRUIT DETECTION METHOD FOR ENTIRE DWARF AND DENSELY PLANTED APPLE TREES BASED ON CHUNKING STRATEGY
/
基于分块策略的矮砧密植苹果整株果树幼果识别方法研究

Yu YANG[1], Xingchen QIAO[1], Huijun ZHAO[2], Yicheng YANG[1], Zhaohui ZHENG[3], Gang ZHANG[4], Xiaole WANG[1], Zhenchao WU*[1]

[1]School of Engineering, Anhui Agricultural University, Hefei, Anhui 230036, China;
[2] Himile Industrial Park, Economic Development Zone, Weifang, Shandong 261500, China
[3] College of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832000, China
[4] School of Intelligent Manufacturing, Anhui Science and Technology University, Chuzhou 239000, China
Corresponding author: Zhenchao Wu
Tel: +8613204655989; E-mail: wuzhenchao@ahau.edu.cn
DOI: https://doi.org/10.35633/inmateh-79-68

Keywords: *Chunking strategy, Apple fruitlets, Branch-leaf occlusion, YOLOv12n, Small-target detection*

ABSTRACT
A chunking strategy based on YOLOv12n has been proposed in this study to address challenges of young fruit detection for entire trees in dwarf and densely planted apple orchards, such as small young fruit volume, high similarity between fruit skin color and surrounding young leaves. This strategy divides an image into sub-images before feeding them into the network for training. Experimental results show that, compared with using only the YOLOv12n model for whole-image recognition, the chunking strategy achieved a precision of 85.57% and a recall of 62.54% in young fruit detection. In contrast, the YOLOv12n model alone, due to ineffective small-target capture and missing edge features, achieved a detection precision of less than 1% for young fruits, making effective detection nearly impossible. The introduction of the chunking strategy achieved a qualitative leap in young fruit detection performance. This strategy provides technical support for precise young fruit counting, subsequent accurate thinning, and growth dynamic monitoring in dwarf and densely planted orchards, and also offers a new implementation approach for small-target detection scenarios in orchards.

摘要
本研究提出了一种基于 YOLOv12n 的分块策略，以解决矮化密植苹果园整树幼果检测的难题，如幼果体积小、果皮颜色与周围幼叶高度相似等。该策略通过将图像分割为子图像后输入网络进行训练。试验结果表明，与仅使用 YOLOv12n 模型进行整图识别，分块策略的使用在幼果识别的准确率上达到 85.57%，召回率达到 62.54%，而单独使用 YOLOv12n 模型则因小目标捕捉失效、边缘特征遗漏，对幼果的检测准确率不足 1%，几乎无法完成有效检测，分块策略的引入实现了幼果识别性能的质的飞跃。该策略为矮砧密植果园幼果精准计数、后续精准疏果及生长动态监测等环节提供了技术支撑，也为果园小目标检测场景提供了新的实现思路。

1. INTRODUCTION
The development of intelligent and precise thinning devices is of great significance for improving the efficiency of the apple industry, reducing production costs, and achieving refined orchard management. Thinning is a key link in apple tree management *(Pawikhum et al., 2025)*. Proper thinning can improve fruit size and quality, maintain tree vigor, and extend the productive lifespan of trees *(Patiyal et al., 2024; Tang et al., 2023)*. Currently, the mainstream thinning methods include manual thinning, chemical thinning, and mechanical thinning. Although manual thinning allows precise control of fruit load, its efficiency is low and labor costs are high *(Du et al., 2024)*. Chemical thinning, while low in cost and simple to operate, is easily affected by climatic conditions, leading to unstable results *(Hussain et al., 2023)*. Mechanical thinning greatly improves efficiency and reduces labor input, but suffers from inaccurate young fruit detection and a tendency to damage fruits *(Hua et al., 2025)*. Therefore, the development of automated intelligent thinning devices is a direction for solving existing problems and promoting industrial upgrading *(Wang et al., 2025)*. The prerequisite for achieving such high-performance devices is the provision of a precise visual perception foundation for apple fruitlets, i.e., reliable detection and localization of young fruits, which directly affects the operational performance and practical value of the device.

Performance improvement of the visual perception module is inseparable from the technical support of deep learning in computer vision fields such as object detection *(Borugadda et al., 2026; Oba et al., 2025)*. Currently, some researchers have applied deep learning networks to fruit detection in complex scenes to enhance the adaptability of the vision module to field environments. *Fu et al. (2020)* improved ZFNet and VGG16 models based on Faster R-CNN for apple detection in dense-foliage fruit trees, achieving average precision improvements to 89.3% and 80.5% for the VGG16 and ZFNet models, respectively, with an average processing time of 0.181 seconds per image for VGG16, indicating almost no impact on processing speed. To quickly and accurately predict the positions of key points such as fruit picking points in complex field environments, *Ma et al. (2025)* proposed a joint detection model for strawberry targets and key points, STRAW-YOLO, based on the YOLOv8-pose network. The improved network achieved a keypoint detection precision of 91.6% and an average precision of 96.0%. *Zheng et al. (2021)* proposed a multi-scale feature fusion network, YOLO-BP, to recognize green citrus, achieving an average precision of 91.55%, but missed detections and false detections occurred in some partially occluded and overlapping fruit scenes. *Liu et al. (2024)* proposed a lightweight algorithm, Faster-YOLO-AP, to detect orchard apples, achieving an average precision of 85.3% for large targets but only 65.6% for small targets. Despite these good results in fruit detection, few studies have captured entire trees for fruit detection in the context of dense small targets such as apple fruitlets in dwarf and densely planted orchards *(Chen et al., 2024)*, to better meet the high precision and high reliability requirements of practical applications.

Compared with deep learning network improvements, the chunking strategy has potential to achieve better small-target detection performance in orchard scene images. Although object detection has developed rapidly, small-target detection remains a major challenge due to the small proportion of pixels occupied by small targets, limited semantic information, susceptibility to complex scene interference, and ease of clustering and occlusion *(Feng et al., 2024; Zhang et al.,2022; Hou et al., 2025)*. The dwarf and densely planted cultivation mode makes small-target detection scenes even more complex, as dense planting causes severe overlap of branches, leaves, and fruits, blurring target boundaries. The color and texture of fruits and leaves converge, and contrast decreases under complex lighting *(Gao et al., 2024)*. However, current small-target detection methods mainly focus on modifying widely used deep learning network architectures *(Wu et al., 2024; Liang et al., 2025)*. Although local performance is improved, effective features are severely attenuated after standard resize operations, and network customization for specific objects limits generalization capability *(Chen et al., 2025)*. In contrast, the chunking strategy preserves the original features of small targets directly through local high-resolution processing, while effectively increasing the number of datasets, avoiding insufficient feature learning during model training due to small dataset sizes, thus opening up a better solution for building a universal detection framework.

The potential of the chunking strategy in young apple fruit detection in large-field-of-view complex scenes is significant, but its practical application still faces multiple key challenges. Currently, limitations of the chunking strategy often lead to missed detections of small targets, such as ignoring dense small fruitlets when chunks are too large, or computational redundancy when chunks are too small *(Dang et al., 2025; Jiang et al., 2023)*. Therefore, it is necessary to determine the optimal chunk size and overlap rate to balance detection accuracy and computational efficiency while avoiding cutting of young fruits. It is also necessary to reduce the destruction of contextual information from the original image during chunking and minimize feature loss caused by local fragmentation. An efficient chunking result recombination strategy needs to be designed to eliminate duplicate detections or missed detection errors at boundaries. These issues have not been fully explored in current immature apple detection research. Therefore, the scientific design and optimization of the chunking strategy is the core prerequisite and implementation path for achieving efficient, high-precision, and large-scale detection of apple fruitlets under the dwarf and densely planted mode.

This study aims to address the key challenges in applying the chunking strategy to young apple fruit detection in dwarf and densely planted orchards. Through scientific chunking strategy design, the problems in young apple fruit detection are solved, and the ability to extract and classify fruit features in complex and diverse environments is enhanced. The chunking strategy is integrated with the YOLOv12n model to achieve accurate detection of dense, occluded apple fruitlets.

2. MATERIALS AND METHODS

This method includes six main processes. First, a mobile phone camera was used to collect images of young apple fruits.

The young apple fruits were annotated to establish a dataset of young apple fruits in dwarf and densely planted orchards. The dataset was divided into training and validation sets through data processing operations. Then, model training was carried out according to three different training strategies, and finally, detection results of young apple fruits in the dwarf and densely planted orchard environment were obtained. An overall schematic diagram of this method is shown in Fig. 1.

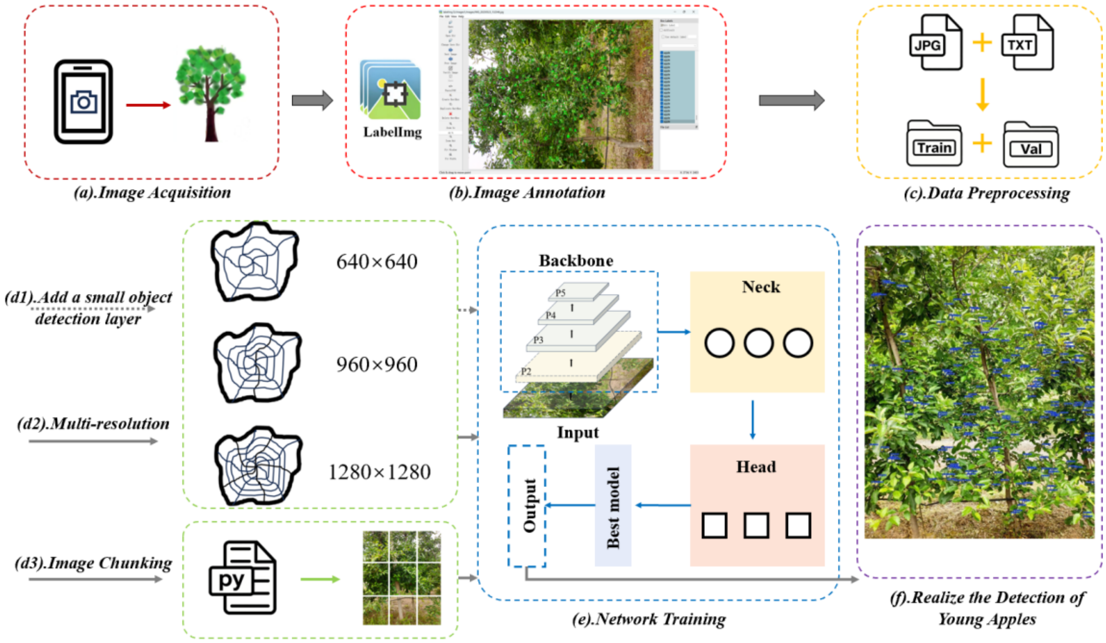


Fig. 1 - Schematic diagram of the overall workflow

2.1 Dataset

2.1.1 Image Acquisition of Young Apple Fruits

On May 16, 2025, images of "Gala" apples were collected at Xinyuan Fruit Industry Group Co., Ltd. in Weihai City, Shandong Province. The planting pattern of the studied apple trees was dwarf and densely planted *(Yuan et al., 2025; Ma et al., 2025)*, without bagging, with a row spacing of 3.5 meters and a plant spacing of 1 meter. The tree age was 7 years, and the length of the tree rows ranged from 30 to 240 meters. A mobile phone camera (Xiaomi Mix 2S) was used to collect images of young apple fruits at a distance of 2.6 meters from the tree row. The image resolution was 4032 × 3024 pixels. A total of 200 original images were collected and saved in JPG format. The acquisition scene is shown in Fig. 2.

Fig. 2 - Apple tree planting pattern and image acquisition method

2.1.2 Image Annotation and Data Preprocessing

Annotation of the young apple fruit dataset was completed using the labeling software LabelImg. The annotation box size was the bounding rectangle of the young fruit. After annotation, the resulting annotation file information was saved as .XML files, which contained key data such as the filename of the young fruit image, the coordinates of the four corners of the annotated rectangular region, and the target classification label. After annotation, the image dataset was split and augmented for training. The 200 images in the dataset were randomly divided into a training set of 160 images and a validation set of 40 images, a ratio of 8:2. After processing with the chunking strategy, the dataset was augmented, expanding the training set from 160 images to 5,600 images. Examples of images in the young apple fruit dataset and the scenes they contain are shown in Fig. 3, including various scenes such as dark lighting, normal lighting, leaf occlusion, and fruit stacking.

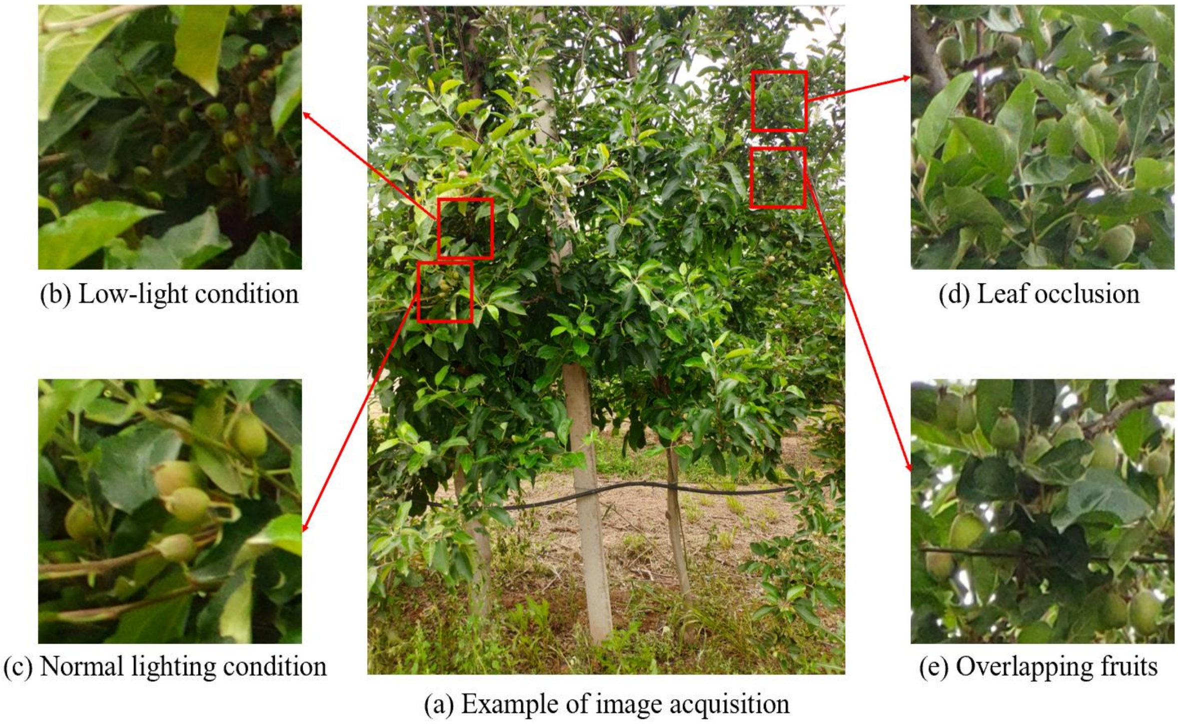


Fig. 3 - Apple samples from a dwarf and densely planted orchard

2.2 Methods

2.2.1 Multi-Resolution Strategy

YOLOv12n is an efficient lightweight model designed for general object detection tasks, with a model structure that includes a small-target detection mechanism suitable for real-time detection scenarios *(Zhang et al., 2025)*. Multiple studies have shown that its innovative region attention mechanism significantly improves the feature extraction efficiency of tiny targets by dynamically dividing feature maps and focusing on local regions. At the same time, the optimized multi-scale feature fusion architecture strengthens the retention of shallow, high-resolution details, effectively solving the problem of information loss of small targets in deep networks. Combined with lightweight design and position-aware enhancement technology, this model has significant advantages in high-precision small-target detection scenarios. Based on its breakthrough performance in balancing accuracy and efficiency, YOLOv12n was used to detect young apple fruits, and the resize parameters were set to 640 × 640, 960 × 960, and 1280 × 1280 to examine their impact on young apple fruit detection.

2.2.2 Adding a Small-Target Detection Layer Strategy

The detection head of YOLOv12n includes three detection layers: P3, P4, and P5 detection heads. Among them, the P3 detection head, which has the largest output feature map, is suitable for detecting small targets due to its smallest global receptive field and high spatial resolution, making it the small-target detection layer *(Wei et al., 2024; Jing et al., 2024)*. Similarly, the P5 detection head with the smallest output feature map is the large-target detection layer, and P4 is the medium-target detection head. Before chunking, the number of small targets was 26,558; after chunking, the number of small targets was 24,315, as shown in Table 1.

Table 1

Statistics of instance counts

Dataset	Images	Instances	Objects		
			Small Objects	Medium Objects	Large Objects
Training Set (Non-chunked)	160	20456	20455	1	0
Validation Set (Non-chunked)	40	6103	6103	0	0
Training Set (Chunked)	5600	22595	18655	3594	346
Validation Set (Chunked)	1400	6685	5660	901	124

Small targets accounted for a large proportion of the overall samples, placing higher demands on the performance of the small-target detection layer. Moreover, due to the small target size of young fruits in the images, their feature information was easily lost during detection, leading to missed detections. In this study, a newly added P2 detection head was set to detect tiny targets approximately 4×4 to 8×8 pixels in size to detect even smaller young apple fruits. The network structure diagram is shown in Fig. 4. Through a series of operations such as convolution, upsampling, and concatenation, feature maps suitable for tiny target detection were generated, thereby improving the detection accuracy of small young fruit targets. Similarly, this strategy was also evaluated at three resolutions, and both strategies were jointly used to comprehensively examine the effect of resolution changes on detection performance.

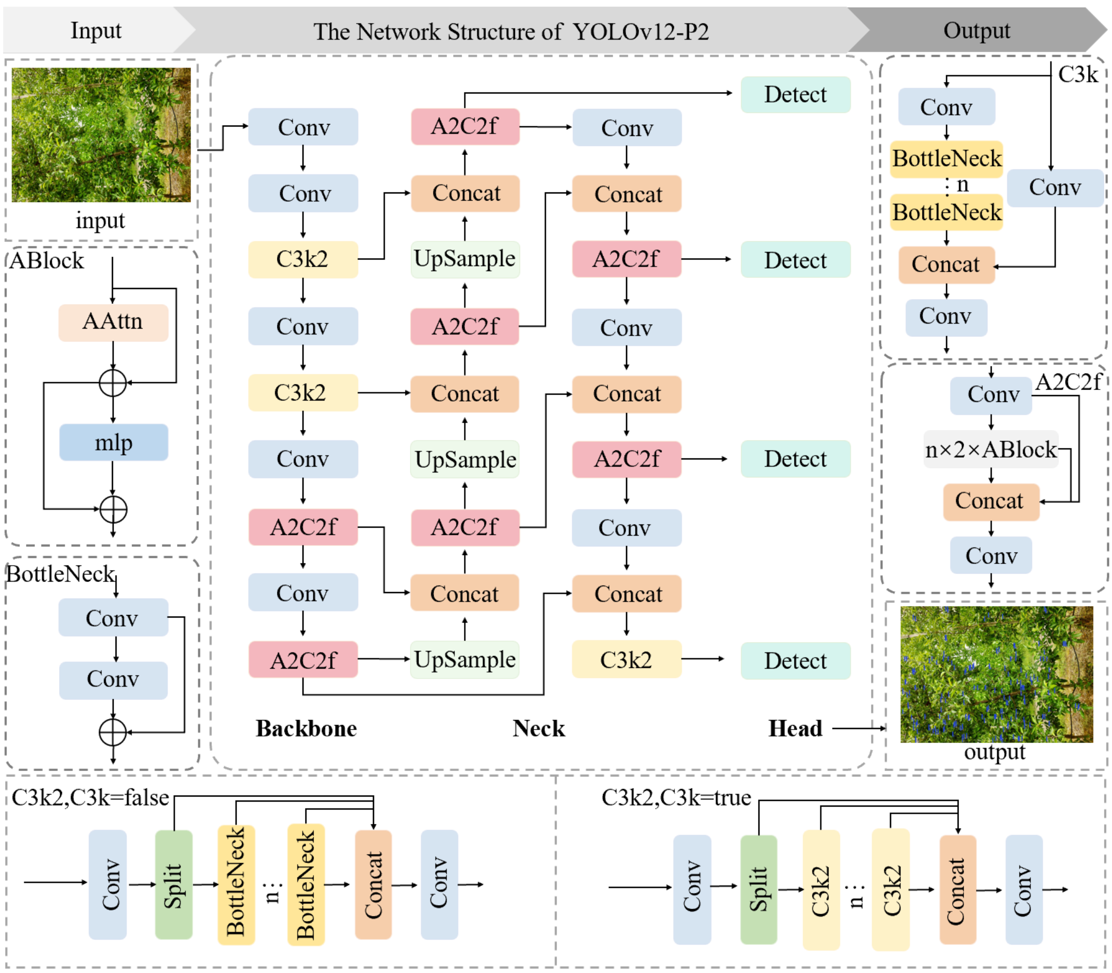


Fig. 4 - Network structure diagram of YOLOv12n-P2

2.2.3 Chunking Strategy

Young apple fruits occupy a very small proportion of pixels in the original image, which poses a challenge to current object detection models based on deep learning networks. Therefore, a chunking strategy was proposed to augment the data and increase the proportion of young apple fruit pixels in the image.

This algorithm started from the upper left corner of the image and progressively moved to the lower right, decomposing the originally large image into multiple small segmented images, each with a pixel size that met the input specifications of the training network. Some segmented images might contain no young apple fruits, while others had an increased proportion of young apple fruits. The traditional chunking method sequentially divides an image based on pixels. When a target crosses chunk boundaries or is unevenly distributed, it easily leads to feature fragmentation and loss of contextual information, affecting detection or counting accuracy. Therefore, a certain overlap was set between adjacent chunks to prevent missing young fruit information at the edges of the images due to chunking. As shown in Fig. 5, three black rectangular boxes represent three adjacent partitions, and the blue area represents the overlapping area between the three partitions.

The overlap length was calculated using formulas (1) and (2).

$$overlap = \frac{x'}{2x - x'} \tag{1}$$

$$overlap = \frac{y'}{2y - y'} \tag{2}$$

where:

$overlap$ is the overlap rate of the chunking strategy, [%]; x is the width of the chunked image on the horizontal axis, [pixel]; y is the height of the chunked image on the vertical axis, [pixel]; x' is the length of the overlapping part on the horizontal axis, [pixel]; y' is the length of the overlapping part on the vertical axis, [pixel].

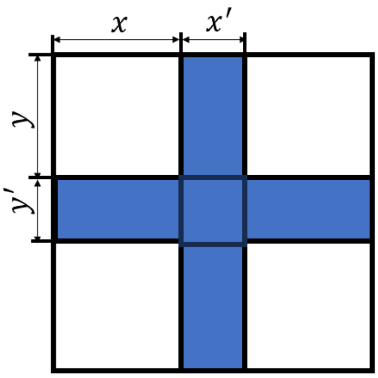


Fig. 5 - Examples of overlap rate setting and overlap region

After the overlap rate was set, the image information was read, and the number of chunks along the horizontal and vertical axes was calculated. The formulas for calculating the number of image chunks required are shown in formulas (3) and (4).

$$a = 1 + \left[\frac{w - x}{x - x'}\right] \tag{3}$$

$$b = 1 + \left[\frac{h - y}{y - y'}\right] \tag{4}$$

where:

a is the number of chunked images required on the horizontal axis; w is the width of the original image, [pixel]; b is the number of chunked images required on the vertical axis; h is the height of the original image, [pixel].

At the same time, to ensure that all young fruit information in the image could be detected, the results were rounded up. However, rounding up caused the relative coordinates of the edge images to exceed the original image during chunking, resulting in blank areas in the chunked images. To avoid this, the relative coordinates were shifted. An example of chunked image shifting is shown in Fig. 6.

The yellow box represents the original calculated partition position of the image, and the red box represents the corrected partition position of the image.

Fig. 6 - Example of chunked image shifting

The segmented small images were set to 640 × 640 pixels in this study to fit the network input size. The chunking strategy was also used to partition the XML annotation files and original label files to avoid re-labeling the label files after image chunking. The relative coordinate calculation formulas are shown in formulas (5) and (6).

$$x'_{\min} = x_{\min} - x_0 \tag{5}$$

$$y'_{\min} = y_{\min} - y_0 \tag{6}$$

where:

$x'_{\min}$ is the relative horizontal coordinate of the bounding box with respect to the chunked image, [pixel]; $y'_{\min}$ is the relative vertical coordinate of the bounding box with respect to the chunked image, [pixel]; $x_{\min}$ is the original horizontal coordinate of the bounding box with respect to the whole image, [pixel]; $y_{\min}$ is the original vertical coordinate of the bounding box with respect to the whole image, [pixel]; x_0 is the horizontal coordinate of the top-left corner of the chunked image, [pixel]; y_0 is the vertical coordinate of the top-left corner of the chunked image, [pixel].

The image to be detected was divided into regions as shown in Fig. 7. The black boxes indicate the chunk positions after setting the overlap rate, the red boxes indicate the chunk positions after shifting, and the blue boxes indicate the chunk positions for the last image. Finally, the detected chunked images were merged to restore their original size.

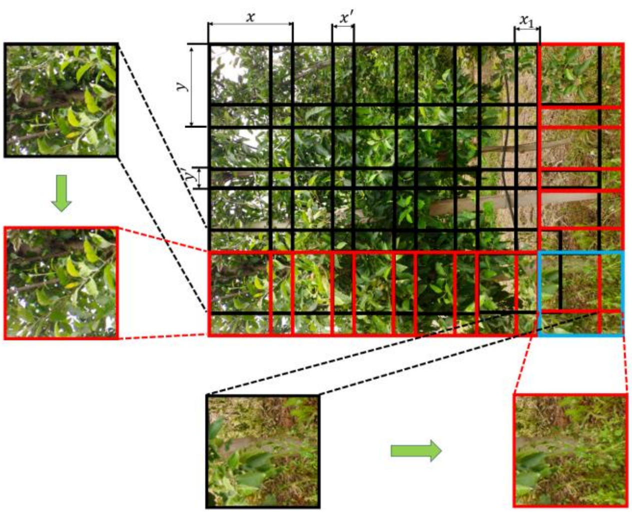


Fig. 7 - Chunk region division of the image to be detected

2.3 Experimental Platform and Training Parameters

The experimental model training platform used an NVIDIA GeForce RTX 4050 graphics card, a 13th Gen Intel(R) Core(TM) i7-13620H processor, and 6 GB of video memory. The network model was trained using the Windows 11 operating system, PyTorch 2.0.0 framework, CUDA 11.8, PyCharm editor, and Python 3.12 programming language. Stochastic gradient descent was used for training, with a momentum of 0.937 and a weight decay of 0.0005. The initial learning rate of the network was set to 0.01, and the number of network training epochs was 300.

2.4 Evaluation Metrics

To evaluate the performance of different strategies in recognizing young apple fruits, the following metrics were used: *Precision (P), Recall (R),* and *Average Precision (AP)*. *AP* is the area under the *P-R* curve and was used as a comprehensive metric reflecting network performance. The calculation formulas for P, R, and AP are shown in formulas (7), (8), and (9).

$$P = \frac{TP}{TP + FP} \times 100\% \tag{7}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{8}$$

$$AP_i = \int_0^1 P_i(R_i)dR_i \tag{9}$$

where: TP (*True Positive*) is the number of correctly detected young apple fruits; *FP (False Positive)* is the number of incorrectly detected young apple fruits; *FN (False Negative)* is the number of young apple fruits that were not detected.

3. RESULTS

3.1 Comparison of Different Detection Models

To systematically evaluate the detection difficulty of this dataset and exclude performance anomalies caused by specific model implementations or training settings, four mainstream lightweight object detection models, YOLOv5n, YOLOv8n, YOLOv11n, and the baseline YOLOv12n, were selected to detect young apple fruits in the dwarf and densely planted orchard environment. The experimental results are shown in Table 2. The P values of YOLOv5n, YOLOv8n, and YOLOv11n on this dataset were only 0.10%, 0.29%, and 0.23%, respectively, and their *P* and *R* values were also below 1%. This result was highly consistent with the baseline performance of YOLOv12n used in this study, indicating that the extremely small size, dense distribution, severe leaf occlusion, and fruit stacking of young apple fruits in this dataset made it nearly impossible for current mainstream object detection models to learn effective young fruit features. These results together confirmed that the baseline model *AP* below 1% was not caused by specific code defects or training errors, but was an objective reflection of the extreme challenge of this dataset.

Table 2

Performance comparison of different detection models on the young apple fruit dataset

Model	Different resolutions	P (%)	R (%)	AP (%)
YOLOv5n	640×640	0.10	0.20	0.05
YOLOv8n	640×640	0.29	0.57	0.17
YOLOv11n	640×640	0.23	0.44	0.36
YOLOv12n	640×640	0.24	0.48	0.12

3.2 Comparison of Three Detection Methods

The performance of three methods in object detection tasks was compared. The experimental results are shown in Table 3. The chunking strategy significantly outperformed the other two methods, achieving an *AP* as high as 77.50%, an *R* of 62.54%, and a *P* of 85.57%, indicating that this method maintained high detection stability and accuracy even in complex scenes. Particularly in environments with many small targets and heavy occlusion, the chunking strategy effectively improved target detection and localization through local enhancement processing. As shown in Fig. 8e, the chunking strategy was the only method among the three that effectively recognized young fruits.

Table 3

Performance comparison of the three detection methods

Strategies	Different resolutions	P (%)	R (%)	AP (%)
YOLOv12n	640×640	0.38	0.75	0.20
	960×960	0.28	0.54	0.24
	1280×1280	0.31	0.61	0.20
YOLOv12n-P2	640×640	0.29	0.57	0.16
	960×960	0.24	0.48	0.13
	1280×1280	0.29	0.57	0.15
YOLOv12n-chunking	640×640	85.57	62.54	77.50

In contrast, the multi-resolution training and the small-target detection layer addition methods performed poorly in this experiment. The highest AP achieved by the multi-resolution training method was only 0.24%, and the highest R was only 0.75%, indicating that this method failed to effectively capture target features, possibly due to insufficient model training in adapting to scale variations. The method of adding a small-target detection layer also failed to significantly improve performance, with the highest AP reaching only 0.16% and the highest R only 0.57%, indicating that simply increasing network depth or the number of detection layers cannot fundamentally solve the problems of semantic loss and feature weakening in small-target detection. Although the multi-resolution training and small-target detection layer addition methods showed slight numerical improvements in P, R, and AP, their extremely low values indicated that these methods were practically unable to detect real targets, with severe missed detections. As shown in Figure 8, methods other than the chunking strategy produced almost no detection boxes.

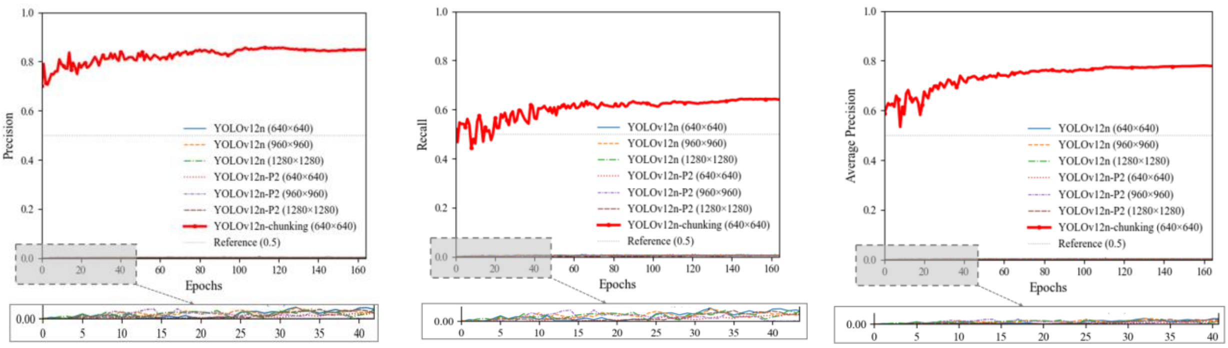
Fig. 8 - Detection effects of the three strategies

The superior performance of the chunking strategy was closely related to the growth characteristics of young apple fruits. Young fruits are typically small in the early stage, highly similar in color to leaves, and often occluded by branches and leaves, making detection extremely challenging *(Liu et al., 2023; Xu et al., 2024)*. When the network needs to process images containing many such tiny, highly similar targets, its receptive field and feature extraction capacity easily reach a bottleneck *(Lin et al., 2025)*. The chunking strategy physically increases the relative size and feature richness of each young fruit in the sub-images by dividing the high-resolution image into smaller sub-images, greatly alleviating the problem of weak small-target feature signals in the original image and enabling the network to capture key discriminative features more clearly. In contrast, although the other two methods enhanced small-target perception to some extent, they did not fundamentally change the proportional relationship between the target and the entire image, thus still suffering from many missed detections and false detections in complex backgrounds.

For small targets such as young apple fruits, a more complex network structure does not necessarily yield better results. Sometimes, a relatively simple and direct method, such as changing the input strategy, can bring more significant performance improvements. This finding is not entirely consistent with the conclusion of *Chutichaimaytar et al. (2025)*, who suggested increasing model depth and complexity to improve small-target detection accuracy. This discrepancy indicates that the optimal detection strategy may vary depending on the specific detection object and its environmental background. For objects with tiny features and complex backgrounds, ensuring that key features are fully represented is more effective than designing complex networks for difficult discrimination.

3.3 Effectiveness Analysis of Each Method

As shown in Figure 9, the performance curves of the three strategies indicate that the model performance achieved a breakthrough improvement after adopting the chunking strategy. The *AP* value significantly increased from less than 1.00% to 77.50%, and *R* and *P* reached 62.54% and 85.57%, respectively. The superior performance of this strategy stemmed from the physical partitioning of the image, which fundamentally increased the relative size of young fruits in the sub-images, allowing the network to capture clearer and richer feature information, thus effectively solving the fundamental problem of extremely weak small-target feature signals *(Wu et al., 2024)*. However, the main cost of this strategy was reduced computational efficiency. Because more sub-images needed to be processed and results fused, the detection time per image was several times that of the end-to-end model. Nevertheless, in offline high-precision analysis scenarios such as young apple fruit yield prediction, trading computation time for unparalleled detection accuracy is a necessary and acceptable trade-off.



(a) (b) (c)

Fig. 9 - Performance curve comparison of the three methods

(a) Comparison of precision curves of the three methods; (b) Comparison of recall curves of the three methods; (c) Comparison of average precision curves of the three methods.

The multi-resolution training strategy achieved AP values below 1% at different resolutions. The intention of this strategy was to enhance the scale invariance of the model, but its failure in young apple fruit detection was attributed to the fact that downsampling operations destroyed the already weak features of young fruits, while upsampling to higher resolutions could not create effective details but instead amplified image noise, interfering with model judgment *(Ma et al., 2022)*. Similar to the detection layer addition strategy, this method was lightweight but did not change the proportional relationship between the target and the entire image, thus having limited effect in addressing the specific difficulties of this study.

As a model enhancement method, the addition of a small-target detection layer aimed to improve small-target perception by introducing a shallower detection head. However, the experimental results did not meet expectations, with the highest AP reaching only 0.16%. The root cause of its failure was that the feature extraction bottleneck occurred in the early stage of the network. The extremely small size of young fruits meant that after several downsampling operations during feature extraction in the backbone network, the effective signal had severely attenuated or been lost. Subsequently, even with the introduction of a detection layer specifically designed for small targets, key information could not be recovered from the already confused feature maps. This strategy had negligible detection effects in the extremely challenging scenario of this study, characterized by extremely small targets and high-background similarity, and was only suitable for general detection tasks with relatively more prominent targets.

In contrast, the failure of the multi-resolution training and small-target detection layer addition strategies highlighted the particularity and challenges of young apple fruit detection. For targets with extremely small features and very low signal-to-noise ratios, traditional methods aimed at enhancing model perception may have limited effect. The most effective strategy is to start from the input side and fundamentally change the relationship between the target and the image scale through physical methods such as chunking, ensuring that the original features of the target are completely captured by the model. The results of this study provide important insights for the detection of similar small targets in agriculture: when the target is small to a certain extent, optimizing the data input strategy should take priority over designing complex models.

3.4 Performance Demonstration and Analysis of the Chunking Strategy

The overall detection effect showed that in real orchard environments with uneven lighting and cluttered backgrounds, the chunking strategy demonstrated excellent overall performance. The model successfully located and recognized tiny young fruit targets with high prediction confidence, as shown in Figure 10. This indicates that the strategy of enhancing local feature representation through chunking processing fundamentally solved the feature loss and receptive field mismatch problems in the baseline method, achieving extremely high recall and precision.

Fig. 10 - Overall detection effect of the chunking strategy and enlarged views of local areas

The enlarged local areas show the model's detailed performance in scenes with partial occlusion, boundary blur, extreme occlusion, and dense targets. While affirming the model's strengths, these examples also reflect its shortcomings and directions for improvement. In Fig.10a, a fruit with approximately 30%–40% leaf occlusion is shown. The model was able to effectively infer and recognize the fruit based on its visible features, providing a reasonable confidence level. This demonstrates the model's ability to complete features and reason under moderate occlusion. However, when the occlusion proportion increases further or the occlusion pattern becomes more complex, the model may still show decreased confidence or missed detections, indicating that its occlusion reasoning mechanism is not yet fully robust. In the future, adversarial occlusion training or structure-aware reasoning modules could be introduced to enhance adaptability to diverse occlusion patterns. In Fig.10b, the chunking strategy significantly improved the distinction between adjacent targets by enhancing local feature resolution. The model generated prediction boxes with accurate spatial positions and clear boundaries in this region, demonstrating excellent discrimination ability for highly overlapping targets. Nevertheless, when target edges are highly fused with the background, the model may still show boundary localization errors, suggesting that the current edge perception mechanism still has room for optimization. Subsequent work could incorporate edge enhancement operators or introduce boundary-aware loss functions to further improve localization accuracy. Fig.10c presents a boundary case of the current method. When a target is severely occluded or presents only very small local features, detection confidence decreases. In the future, introducing attention-based contextual reasoning modules *(Xu et al., 2023)* could allow the model to better utilize surrounding environmental information to infer the presence of extremely occluded targets. Fig.10d shows the model's excellent ability to distinguish dense targets. For closely clustered young fruit groups, the model did not incorrectly merge them into a single detection target but generated multiple independent and accurately bounded prediction boxes. However, when target spacing is extremely small and targets are heavily adhered to each other, the model may still suffer from duplicate detections or missed detections, indicating that its instance separation capability in extremely dense scenes needs strengthening. This demonstrates that the high-resolution local contextual information provided by the chunking strategy is crucial for distinguishing spatially adjacent targets with similar appearances.

4. CONCLUSIONS

This study addressed the problems of extremely small young apple fruit targets and high detection difficulty in dwarf and densely planted orchard environments by proposing a detection method based on the YOLOv12n-chunking strategy. Compared with methods such as multi-resolution training and adding a small-target detection layer, the chunking strategy adopted in this study achieved significant improvements in detection performance, with an *AP* of 77.50%, while the *AP* values of the multi-resolution training and small-target detection layer addition methods were both below 1%. The chunking strategy enhanced local feature expression through image preprocessing, effectively overcoming the detection bottlenecks caused by extremely small target scales and complex backgrounds, providing an effective solution for small-target agricultural detection in large-field-of-view orchard images. However, the current detection accuracy of 77.50% still has room for improvement. In the future, the dataset will be expanded with multi-season, multi-variety, and multi-weather-condition images, and lightweight design and advanced algorithms will be integrated to further improve the model's generalization ability and deployment suitability.

5. ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China Youth Program (32501768); Anhui Provincial Natural Science Foundation of Youth Program (2508085QC086), Scientific Research Project of Anhui Education Department (2025AHGXZK40673), Scientific Research Start-up Foundation of Anhui Agricultural University (rc412405).

6. REFERENCES

[1] Borugadda P., Kalluri H.K., (2026). A comprehensive analysis of artificial intelligence, machine learning, deep learning and computer vision in food science, *Journal of Future Foods*, vol.6, no.6, pp.971-987, Beijing/China. DOI: https://doi.org/10.1016/j.jfutfo.2025.07.002

[2] Chen L.-H., Bampis C.G., Li Z., Krasula L., Bovik A.C., (2025). Estimating the resize parameter in end-to-end learned image compression, *Signal Processing: Image Communication*, vol.135, article no.117277, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.image.2025.117277

[3] Chen Y., Niu Y., Cheng W., Zheng L., Sun D., (2024). Apple detection method in the natural environment based on improved YOLOv5 (自然环境下基于改进 YOLOv5 的苹果检测方法), *INMATEH Agricultural Engineering*, vol.72, no.1, pp.183-192, Bucharest/Romania. DOI: https://doi.org/10.35633/inmateh-72-17

[4] Chutichaimaytar P., Zhang Z., Kaewtrakulpong K., Ahamed T., (2025). An improved small object detection CTB-YOLO model for early detection of tip-burn and powdery mildew symptoms in coriander (Coriandrum sativum) for indoor environment using an edge device, *Smart Agricultural Technology*, vol.12, article no.101142, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.atech.2025.101142

[5] Dang H., He L., Shi Y., Janneh L.L., Liu X., Chen C., Li R., Ye H., Chen J., Majeed Y., Kou X., Fu L., (2025). Growth characteristics based multi-class kiwifruit bud detection with overlap-partitioning algorithm for robotic thinning, *Computers and Electronics in Agriculture*, vol.229, article no.109715, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.109715

[6] Du Y., Han Y., Su Y., Wang J., (2024). A lightweight model based on you only look once for pomegranate before fruit thinning in complex environment, *Engineering Applications of Artificial Intelligence*, vol.137, Part A, article no. 109123, Amsterdam / Netherlands. DOI: https://doi.org/10.1016/j.engappai.2024.109123

[7] Feng L., Liu Y., Yang H., Jia Z., Guan J., Zhu H., Hou Y., (2024). Detection of apple leaf diseases target based on improved YOLOv7 (基于改进 YOLOv7 的苹果叶片病害目标检测), *INMATEH Agricultural Engineering*, vol.72, no.1, pp.280-290, Bucharest/Romania. DOI: https://doi.org/10.35633/inmateh-72-26

[8] Fu L., Majeed Y., Zhang X., Karkee M., Zhang Q., (2020). Faster R–CNN–based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting, *Biosystems Engineering*, vol.197, pp.245-256, London/United Kingdom. DOI: https://doi.org/10.1016/j.biosystemseng.2020.07.007

[9] Gao A., Tian Z., Ma W., Song Y., Ren L., Feng Y., Qian J., Xu L. (2024). Fruits hidden by green: an improved YOLOV8n for detection of young citrus in lush citrus trees, *Frontiers in Plant Science*, vol.15, article no.1375118, Lausanne/Switzerland. DOI: https://doi.org/10.3389/fpls.2024.1375118

[10] Hou S., Pang Y., Wang J., Hou J., Wang B., (2025). RS-YOLO: A highly accurate real-time detection model for small-target pest, *Smart Agricultural Technology*, vol.12, article no.101212, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.atech.2025.101212

[11] Hua W., Zhang Z., Zhang W., Liu X., Hu C., He Y., Mhamed M., Li X., Dong H., Saha C.K., Khan W.U., Abid F., Abdelhamid M.A., (2025). Key technologies in apple harvesting robot for standardized orchards: A comprehensive review of innovations, challenges, and future directions, *Computers and Electronics in Agriculture*, vol.235, article no.110343, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2025.110343

[12] Hussain M., He L., Schupp J., Lyons D., Heinemann P., (2023). Green fruit segmentation and orientation estimation for robotic green fruit thinning of apples, *Computers and Electronics in Agriculture*, vol.207, article no.107734, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.107734

[13] Jiang H., Sun X., Fang W., Fu L., Li R., Cheein F.A., Majeed Y., (2023). Thin wire segmentation and reconstruction based on a novel image overlap-partitioning and stitching algorithm in apple fruiting wall architecture for robotic picking, *Computers and Electronics in Agriculture*, vol.209, article no.107840, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.107840

[14] Jing R., Zhang W., Liu Y., Li W., Li Y., Liu C., (2024). An effective method for small object detection in low-resolution images, *Engineering Applications of Artificial Intelligence*, vol.127, Part A, article no.107206, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.engappai.2023.107206

[15] Liang X., Xiang J., Qin S., Xiao Y., Chen L., Zou D., Ma H., Huang D., Huang Y., Wei W., (2025). Small target detection algorithm based on SAHI-improved-YOLOv8 for UAV imagery: A case study of tree pit detection, *Smart Agricultural Technology*, vol.12, article no.101181, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.atech.2025.101181

[16] Lin C., Jiang W., Zhao W., Zou L., Xue Z., (2025). DPD-YOLO: dense pineapple fruit target detection algorithm in complex environments based on YOLOv8 combined with attention mechanism, *Frontiers in Plant Science*, vol.16, article no.1523552, Lausanne/Switzerland. DOI: https://doi.org/10.3389/fpls.2025.1523552

[17] Liu P., Yin H., (2023). YOLOv7-Peach: An algorithm for immature small yellow peaches detection in complex natural environments, *Sensors*, vol.23, no.11, article no.5096, Basel/Switzerland. DOI: https://doi.org/10.3390/s23115096

[18] Liu Z., Abeyrathna R.M.R.D., Sampurno R.M., Nakaguchi V.M., Ahamed T., (2024). Faster-YOLO-AP: A lightweight apple detection algorithm based on improved YOLOv8 with a new efficient PDWConv in orchard, *Computers and Electronics in Agriculture*, vol.223, article no.109118, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.109118

[19] Ma J., Han X., Zhang X., Li Z., (2022). MLWAN: multi-scale learning wavelet attention module network for image super resolution, *Sensors*, vol.22, no.23, article no.9110, Basel/Switzerland. DOI: https://doi.org/10.3390/s22239110

[20] Ma W., Ma Z., Hou Y., Lu S., Liang G., Mao J., Chen B., (2025). The 'green code' in fruit cultivation: research progress and prospects of rootstock control tree dwarfing, *Plant, Cell & Environment*, vol.48, no.12, pp.8860-8872, Oxford/United Kingdom. DOI: https://doi.org/10.1111/pce.70151

[21] Ma Z., Dong N., Gu J., Cheng H., Meng Z., Du X., (2025). STRAW-YOLO: A detection method for strawberry fruits targets and key points, *Computers and Electronics in Agriculture*, vol.230, article no.109853, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2024.109853

[22] Oba Y., Doi H., (2025). Accelerating ecosystem monitoring through computer vision with deep metric learning, *Ecological Complexity*, vol.62, article no.101124, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.ecocom.2025.101124

[23] Patiyal V., Rana V.S., Rana N., Hashem A., Abd Allah E.F., Sharma S., (2024). Appraisal of extended cane length and fruit thinning strategies on the performance of growth yield and quality of kiwifruit, *Heliyon*, vol.10, no.8, article no.e29546, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.heliyon.2024.e29546

[24] Pawikhum K., Yang Y., He L., Heinemann P., (2025). Development of a machine vision system for apple bud thinning in precision crop load management, *Computers and Electronics in Agriculture*, vol.236, article no.110479, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2025.110479

[25] Tang Y., Qiu J., Zhang Y., Wu D., Cao Y., Zhao K., Zhu L., (2023). Optimization strategies of fruit detection to overcome the challenge of unstructured background in field orchard environment: a review, *Precision Agriculture*, vol.24, no.4, pp.1183-1219, Dordrecht/Netherlands. DOI: https://doi.org/10.1007/s11119-023-10009-9

[26] Wang X., Yang T., Chen Z., Liu J., He Y., (2025). NLDETR-YOLO: A decision-making method for apple thinning period, *Scientia Horticulturae*, vol.341, article no.113991, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.scienta.2025.113991

[27] Wei W., Cheng Y., He J., Zhu X., (2024). A review of small object detection based on deep learning, *Neural Computing and Applications*, vol.36, no.12, pp.6283-6303, London/United Kingdom. DOI: https://doi.org/10.1007/s00521-024-09422-6

[28] Wu S., Lu X., Guo C., Guo H., (2024). Accurate UAV small object detection based on HRFPN and EfficentVMamba, *Sensors*, vol.24, no.15, article no.4966, Basel/Switzerland. DOI: https://doi.org/10.3390/s24154966

[29] Wu Z., Wang X., Jia M., Liu M., Sun C., Wu C., Wang J., (2024). Dense object detection methods in RAW UAV imagery based on YOLOv8, *Scientific Reports*, vol.14, no.1, article no.18019, London/United Kingdom. DOI: https://doi.org/10.1038/s41598-024-69106-y

[30] Xu D., Xiong H., Liao Y., Wang H., Yuan Z., Yin H., (2024). EMA-YOLO: A novel target-detection algorithm for immature yellow peach based on YOLOv8, *Sensors*, vol.24, no.12, article no.3783, Basel/Switzerland. DOI: https://doi.org/10.3390/s24123783

[31] Xu W., Xu T., Thomasson J.A., Chen W., Karthikeyan R., Tian G., Shi Y., Ji C., Su Q., (2023). A lightweight SSV2-YOLO based model for detection of sugarcane aphids in unstructured natural environments, *Computers and Electronics in Agriculture*, vol.211, article no.107961, Amsterdam/Netherlands. DOI: https://doi.org/10.1016/j.compag.2023.107961

[32] Yuan B., Zhang H., Li Y., Cao X., Sun L., Jing L., Xue L., Liu C., Fan G., Wang J., (2025). Design and experimental evaluation of a two-stage domain-segmented harvesting device for densely planted dwarf apple orchards, *AgriEngineering*, vol.7, no.5, article no.135, Basel/Switzerland. DOI: https://doi.org/10.3390/agriengineering7050135

[33] Zhang C., Kang F., Wang Y., (2022). An improved apple object detection method based on lightweight YOLOv4 in complex backgrounds, *Remote Sensing*, vol.14, no.17, article no.4150, Basel/Switzerland. DOI: https://doi.org/10.3390/rs14174150

[34] Zhang J., Zhang C., Xu L., Mao B., Chen Z., (2025). DBA-YOLO: a mamba-incorporated dual-branch axial feature enhancement network for small object detection in traffic scenarios, *Signal, Image and Video Processing*, vol.19, article no.991, London/United Kingdom. DOI: https://doi.org/10.1007/s11760-025-04572-z

[35] Zheng Z., Xiong J., Lin H., Han Y., Sun B., Xie Z., Yang Z., Wang C., (2021). A method of green citrus detection in natural environments using a deep convolutional neural network, *Frontiers in Plant Science*, vol.12, article no.705737, Lausanne/Switzerland. DOI: https://doi.org/10.3389/fpls.2021.705737