# APPLE DETECTION METHOD IN THE NATURAL ENVIRONMENT BASED ON IMPROVED YOLOv5

/

# 基于改进 YOLOv5 的自然环境下苹果检测方法

**Yongpeng CHEN, Yi NIU, Weidong CHENG*), Laining ZHENG, Dongchao SUN** [1]
School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255091, China
*Tel: +86-13589596562; E-mail: 45035828 @qq.com*
*DOI: https://doi.org/10.35633/inmateh-72-17*

***Keywords:*** *Apple; target detection; YOLOv5; natural environment; deep learning*

## ABSTRACT

*To improve the accuracy of apple fruit recognition, enhance the efficiency of automatic picking robots in orchards, and provide effective visual guidance for the picking robot, a target recognition network model based on improved YOLOv5 is proposed. Firstly, the original apple images collected and the data images obtained by different data enhancement methods are used to establish a dataset of 1,879 images, and the dataset is divided into the training set and the test set under 8:2; then for the problem of low detection accuracy of apple fruits in the natural environment due to the mutual obstruction of apple fruits, this paper modifies the backbone network of YOLOv5 by adding the attention mechanism of the Transformer module, the Neck structure is changed from the original PAFPN to BiFPN that can perform two-way weighted fusion, and the Head structure adds the P2 module for shallow down sampling; finally, the recognition test is performed on the dataset, and a comparative analysis is performed according to different evaluation indexes to verify the superiority of the proposed model. The experimental results show that: compared with other existing models and the single-structure improved YOLOv5 model, the comprehensive improved model proposed in this paper has higher detection accuracy, resulting in an increase of 3.7% in accuracy.*

## 摘要

*为提高对苹果果实识别的准确率，提升果园自动采摘机器人的工作效率，为给采摘机械手提供有效的视觉引导，提出了一种基于改进 YOLOv5 目标识别网络模型。首先使用采集到的苹果原始图像以及其搭配不同数据增强方式得到的数据图像共 1879 幅建立数据集，按照 8:2 将数据集划分成训练集与测试集；然后针对苹果果实之间相互遮挡导致自然环境下苹果果实检测精度低的问题，本文将 YOLOv5 的骨干网络进行改动，增添具有注意力机制的 Transformer 模块，Neck 结构由原来的 PAFPN 改成可以进行双向加权融合的 BiFPN，Head 结构增加了浅层下采样的 P2 模块；最后，对数据集进行识别测试,并根据不同评价指标进行对比分析,验证所建模型的优越性。实验结果表明：相比于其他已有模型以及单一结构改进后的 YOLOv5 模型，本文提出的综合改进模型具有更高的检测精度，使识别精确率提升了 3.7%。*

## INTRODUCTION

The orchard environment is more complex, branch and leaf shading, fruit overlap, light changes, etc. will affect the detection accuracy of the model, resulting in misdetection, leakage, and other problems; in addition, due to the limited arithmetic resources of the embedded platform carried by the picking robot, the detection speed of the complex model cannot meet the task real-time demand, and it is difficult to deploy. Improving the detection speed while ensuring the detection accuracy becomes the main difficult problem and research point of Apple detection in an unstructured environment.

The actual harvest situation faced by the picking robot in the apple orchard is shown in Figure 1. This robot can pick apples that are not obstructed or only covered by leaves. However, apples that are obstructed by branches or other fruits cannot be harvested by the picking robot, because in these situations, if apples are picked directly without accurate recognition, the robot's apples, grasping end effectors, and mechanical picking arms may be damaged, leading to the failure of the picking operation. Therefore, it is crucial for harvesting robots to automatically recognize apple targets that can be grabbed.

---

[1] *Yongpeng Chen,M.S. Stud. Eng.;Weidong Cheng, As. Ph.D. Eng.; Niu Yi,M.S. Stud. Eng.; Laining Zheng, M.S. Stud. Eng.; Dongchao Sun, M.S. Stud. Eng.*

**Fig. 1 - Five-degree-of-freedom apple picking robot**

With the development of artificial intelligence, artificial neural networks have been widely used in many research fields in recent years. For example, in the field of economy, stacking and deep neural network models are deployed separately on feature engineered and bootstrapped samples for estimating trends in prices of underlying stocks during pre- and post-COVID-19 period (*Ghosh I. et al., 2020*). Grey relational analysis (GRA) and artificial neural network model are used for the prediction of consumer exchange traded funds (ETFs) (*Malinda M. et al., 2020*). In the field of industry, a relatively simple network control system solution based on fuzzy logic is proposed by using the relevant ideas of neural network (*Precup R. et al., 2020*). A prediction model of maximum power generation of photovoltaic modules based on fuzzy logic principle and artificial neural network is developed (*Mirko S. et al., 2019*). In the field of agriculture, a new deep learning architecture, called VddNet (Vine Disease Detection Net, is proposed for the detection of grape diseases (*Kerkech M. et al., 2020*). Using the combination of machine vision and deep learning, the early Fusarium wilt disease was identified in real time in the potato production system (*Afzaal, H. et al., 2021*).

Up till now, many deep learning network architectures, such as YOLOv2 (*Bresilla K. et al., 2019*), YOLOv3 (*Wu X. et al., 2020, Zhao H. et al., 2021*), Faster R-CNN (*Gao F. et al., 2020, Gene-Mola J. et al., 2019*), DaSNet v2 (*Wang D. et al., 2019*), R-FCN (*Kang H. et al., 2020*), LedNet (*Kang H. et al., 2021*), FCOS (*Long Y. et al., 2021*), DaSNet (*Kang H. et al., 2019*), and Mask R-CNN (*Gene-Mola J. et al., 2020*), have been successfully applied in apple target recognition based on deep learning, and have been successfully applied in the detection of fruit targets on apple trees. The relevant study status is shown in Table 1.

**Table 1**

**Research on apple target recognition, based on deep learning technology**

| Networks Model | Precision (%) | Recall (%) | mAP (%) | Average Detection Speed (s/pic) |
|---|---|---|---|---|
| YOLOv2 | — | — | 90 | 0.333 |
| YOLOv3 | 97 | 90 | 87.71 | 0.016 |
| Faster R-CNN | — | — | 89.3 | 0.181 |
| DaSNet-v2 | 87.3 | 86.8 | 88 | 0.437 |
| R-FCN | 95.1 | 85.7 | | 0.187 |
| LedNet | 85.3 | 82.1 | 82.6 | 0.028 |
| FCOS | 86.3 | 84.2 | 84.5 | 0.256 |
| DaSNet | 87.3 | 86.8 | 88 | 0.437 |
| Mask R-CNN | 85.7 | 90.6 | — | — |

However, throughout the studies of apple target recognition based on deep learning, although the recognition accuracy of most existing apple detection models was high, the real-time performance of many of them was insufficient, due to its high complexity, large number of parameters, and large size.

In this study, the apple tree fruit is used as the research object. A lightweight apple target real-time recognition algorithm based on improved YOLOv5 for picking robots is proposed, and the goal of this algorithm is to realize the real-time recognition of fruits in different branch shading situations on apple trees and to reduce the loss of apple picking.

## MATERIALS AND METHODS

### Data set preparation

The image data used in this experiment are obtained from the field shooting in the orchard, the location is the apple orchard in Beibuhou Village, Malianzhuang Town, Laixi City, Shandong Province. The equipment is a Hikvision color industrial camera, and the shooting distance from the apple is 20-80 cm. The shooting angles are the left side, the right side, the top side, the back side, and the front side with five directions. The image saving format is JPEG, and the resolution is 2400×1600. The color of the apples is slightly different depending on the natural light intensity, with the lighted fruits showing bright red and the backlit fruits showing dark red. To adapt to different working environments, image data acquisition was carried out in the morning, afternoon, and evening, and a total of 1281 original Apple images were collected. Including strong light, low light, and artificial supplemental light fruit images, etc. Figure 2 shows some of the acquired images.

**Fig. 2 - Some of the collected apple images**

### Image data preprocessing

The rectangle function in Labelling is used to label the apple fruits in the image. For the marking process, the completely exposed apples are marked on the inside of the rectangle, the exposed parts of apple fruits that are obstructed or placed together are marked in the rectangle, and the apple fruits with less than 10% unobstructed parts appearing at the image boundary will not be marked. Therefore, the .xml file is obtained. Then the dataset is divided in an 8:2 ratio, with 80% being the training set and the remaining 20% being the testing set. The final number of samples for the training and testing sets is 1632 and 207, respectively.

### Introduction to YOLO v5

YOLO v5 is a single-stage target detection algorithm, which adds some new improvement ideas based on YOLO v4 so that its speed and accuracy have greatly improved performance. The main improvement ideas are shown below:

- Input side: In the model training stage, some improvement ideas are proposed, mainly including Mosaic data enhancement, adaptive anchor frame calculation, and adaptive image scaling;
- Benchmark network: fusing some new ideas from other detection algorithms, mainly including the Focus structure and the CSP structure;
- Neck Networks: target detection networks tend to insert layers between the Backbone and the final Head output layer. The FPN+PAN structure was added in Yolov5 ;
- Head Output Layer: the anchor frame mechanism of the output layer is the same as YOLO v4. The main improvements are the loss function GIOU_Loss for training, and DIOU_nms for prediction frame screening.

The original Yolov5 network architecture is as follows:

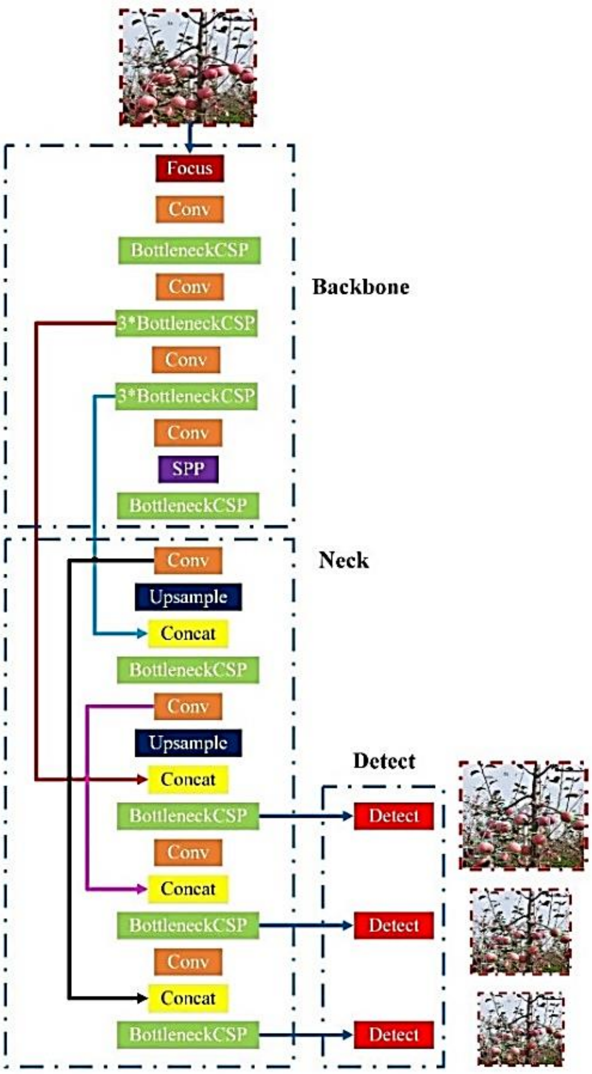



**Fig. 3 - The architecture of the original YOLOv5s network**

## Backbone network optimization

Firstly, the Yolo v5 model backbone network is improved, the biggest drawback of the convolution module is that the convolution parameters need to be set manually, such as the number of convolution kernels (i.e., the number of output channels), the convolution kernel size, the step size, the number of groups, and so on.

In this study, the CSPDDarknet53 backbone network of Yolo v5 is optimized and improved by drawing on the Transform idea. Figure 4 shows the original backbone network of the Yolo v5 model as well as the schematic structure of the improved backbone network.

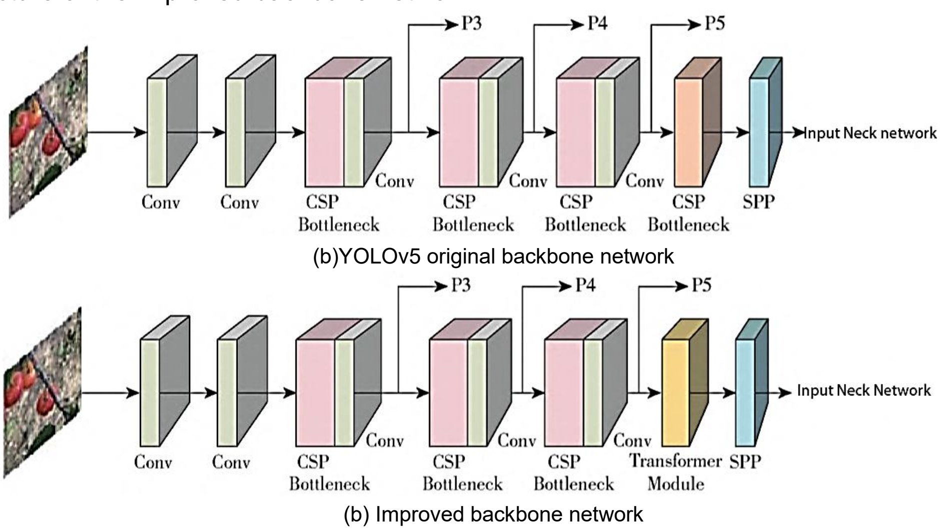



**Fig. 4 - Comparison between improved backbone network and original network**

As can be seen from Fig.3, the improved network replaces the pre-CSPD-Darknet module of the pooled pyramid network used for sensory field feature fusion with the Transformer module, whose outputs are fed into the next level of the Neck network together with the P3, P4, and P5 feature outputs.

Compared to the CSP Bottleneck module, the Transformer module has an attentional mechanism that allows focused extraction of global features. This point can help the detection system to consciously extract the feature points belonging to the apples, to realize better detection effect as well as higher detection accuracy.

**Neck Network Optimization**

The proposal of PANet proves the effectiveness of bidirectional fusion *(Zhao Y. et al., 2016)*, while the bidirectional fusion of PANet is simpler, this paper introduces the BiFPN structure to improve it, as shown in Fig. 5.

For the current research on apple recognition, the following problems are mainly solved by using BiFPN structure as a Neck network for feature fusion:

(1) Apple Small Target Detection Problem: BiFPN, as a modular repeating feature network layer, can obtain a more advanced feature fusion approach compared to PAFPN, increasing the coupling of features at various scales, especially the shallow features that contribute to small target detection.

(2) Apple target overlap problem: Since BiFPN adopts cross-scale connectivity, different detection features can be suppressed or feature expression can be enhanced according to cross-scale weights, thus alleviating the recognition inaccuracy caused by detection target overlap.

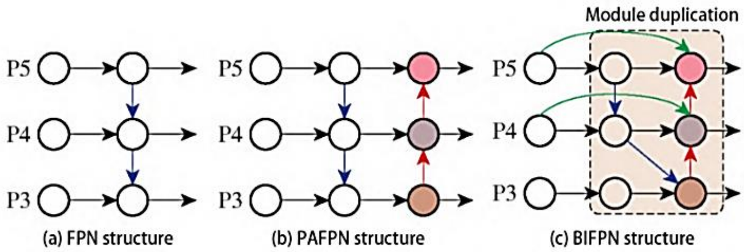



**Fig. 5 - Comparison diagram of different neck network structures**

**Head network optimization**

In the Yolo v5 algorithm, the input features of the Head network*(Xia Y et al., 2013)* are selected as the convolution outputs with a down sampling of 32, 16, and 8, i.e., corresponding to P5, P4, and P3 in Fig. The larger the down sampling, the deeper the convolution, and the stronger the semantic expression of the features, which is better for classification in general, but loses more positional information, while the shallower features with smaller down sampling contain more positional information, which is beneficial for the position detection of small objects *(Fan C. et al., 2020)*. According to this feature of the neural network, this paper optimizes the Head output network as follows: the shallow feature output P2 (4 times down sampling rate) is used as an input feature of the Neck network, which jointly fuses the P3, P4, and P5 features, and the Neck network outputs four features with different scales *(Zhang Q. et al., 2021)* as inputs to the Head branch, to improve the position detection accuracy of small targets. The optimization results are shown in Fig. 6.

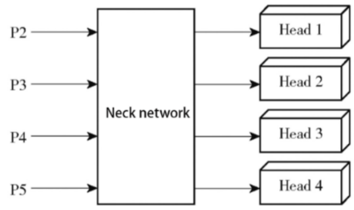



**Fig. 6 - Head network optimization**

**Network Loss Function Optimization**

The loss function L of the YOLOv5 network mainly consists of regression frame prediction loss Lloc, confidence loss $Lconf$, and target classification loss $Lclass$, which is calculated *(Yan B. et al., 2021, Yan B. et al., 2022)*.

$$L = L_{conf} + L_{class} + L_{loc} \tag{1}$$

Among them, the confidence loss and the target classification loss are calculated using the binary cross-entropy loss function (BCELoss), calculated as:

$$BCELoss = \begin{cases} -logp' & (y=1) \\ -log(1-p') & (y=0) \end{cases} \tag{2}$$

where $BCELoss$ denotes the $BCE$ loss function, $p'$ denotes the predicted value of the sample, y denotes the true category of the sample, $y=1$ means it belongs to the target of the category, and $y=0$ means it does not belong to the target of the category.

The GIoU Loss function is used for the regression box prediction loss. The GIoU loss not only takes into account the overlapping region between the real box and the predicted box but also pays attention to the non-overlapping region. This better reflects the distance between the two frames, and thus the regression of the target frame will be more stable, avoiding the target frame regression dispersion problem that occurs when using IOU for model training.

**Network training**

Based on a Lenovo Legion Y7000P computer (Intel(R) Core(TM) i5-11400H CPU, 2.6GHz, 16GB RAM; NVIDIA GeForce RTX 3050 GPU, 6GB video memory), the Pytorch deep learning framework was built under the Windows 10 operating system, using the Python language to write the program and call the required libraries, such as CUDA, Cudnn and OpenCV, to achieve the training and testing of the fruit picking method recognition model for picking robots.

A stochastic gradient descent (SGD) method was used to train the modified YOLOv5 network in an end-to-end joint approach. Four samples were used as a batch size for model training, and the BN layer was used for regularization each time the weights were updated, with Momentum set to 0.937, Decay set to 0.000 5, Initial Learning Rate (ILP) set to 0.01, IOU threshold set to 0.01, and Hue, Saturation, and Brightness set to 0.01. The model was trained using the stochastic gradient descent (SGD) method in an end-to-end joint approach. (H), saturation (S) and brightness (V) enhancement factors were set to 0.015, 0.7, and 0.4, respectively, and a total of 300 rounds of training (Epochs) were performed. The obtained weight files are saved after the model training and the performance of the recognition model is evaluated on the test set. After eliminating a large number of redundant prediction frames by post-processing operations such as Non-maximum suppression (NMS), the final output of the network is the prediction category of the apple picking method with the highest confidence score and the coordinates of the fruit location prediction frames are returned.

From the training results, it can be seen that with the increase in the number of training times, the precision and recall as well as the mAP of the improved YOLO v5s network model are relatively significantly improved compared with the YOLO v5s network model. It can be seen that the network model proposed in this paper can effectively carry out the study of multi-target detection problems and meet the requirements of fruit image recognition by picking robots.

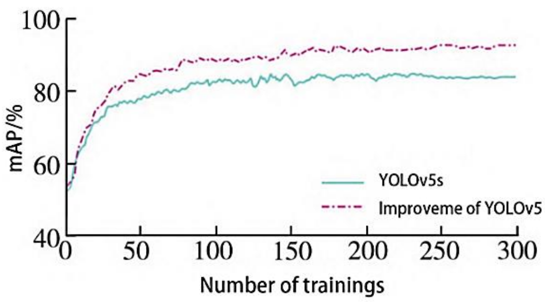



**Fig. 7 - mAP curve**

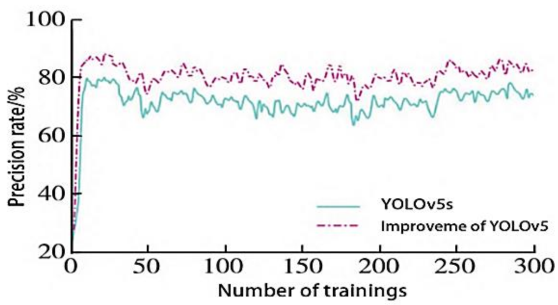



**Fig. 8 - Precision rate curve**

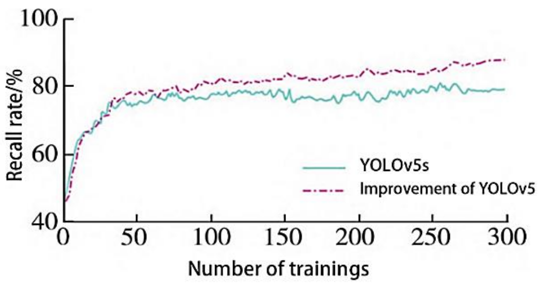


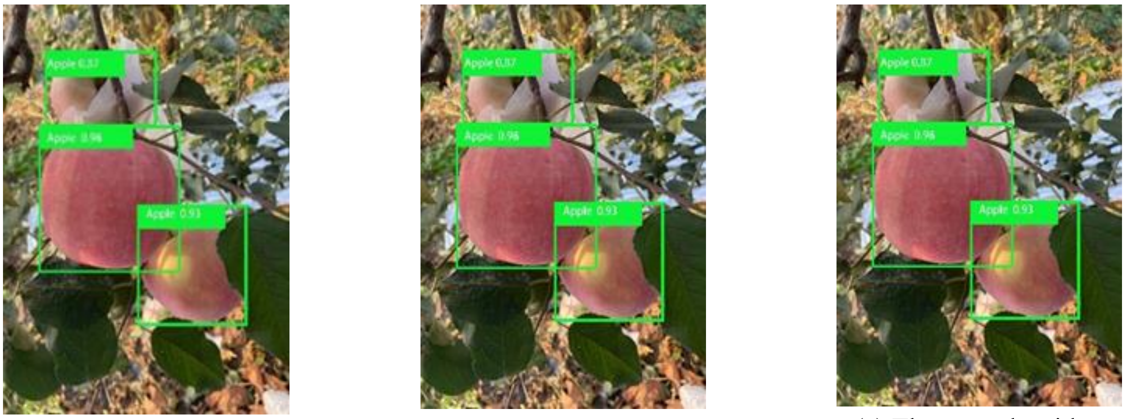

**Fig. 9 - Recall rate curve**

## RESULTS

### Target identification result

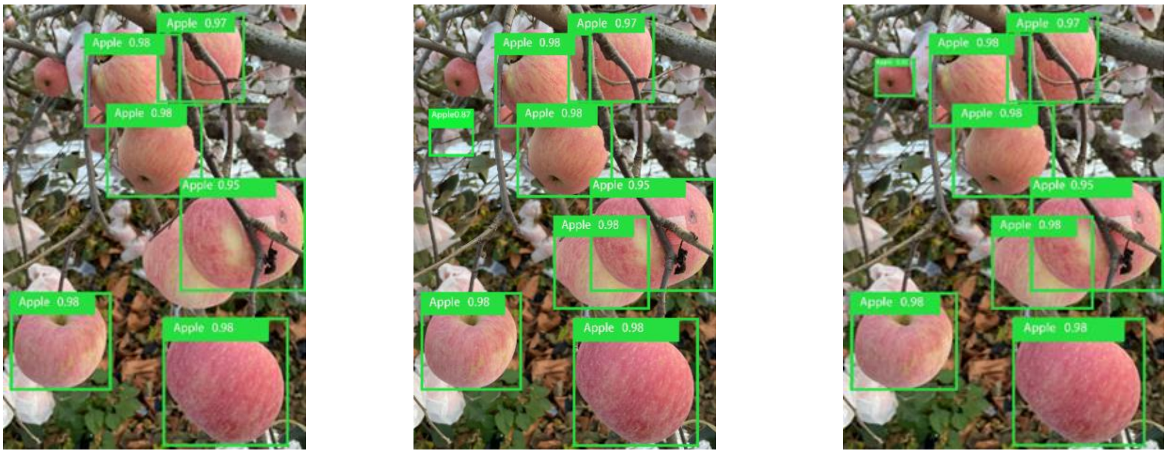
To verify the performance of the fruit, this study further analyzed the recognition results of the model for 200 test set images based on the real-time recognition model of the apple picking robot with improved design of YOLOv5s. In the 200 test set images, there are 2336 apple targets, of which the number of fruit-grabbable targets is 1007 and the number of non-fruit-grabbable targets is 1329.

Examples of the recognition results of the three network models are shown in Fig. 10, which are the recognition results of the apple target under cloudy and sunny conditions.



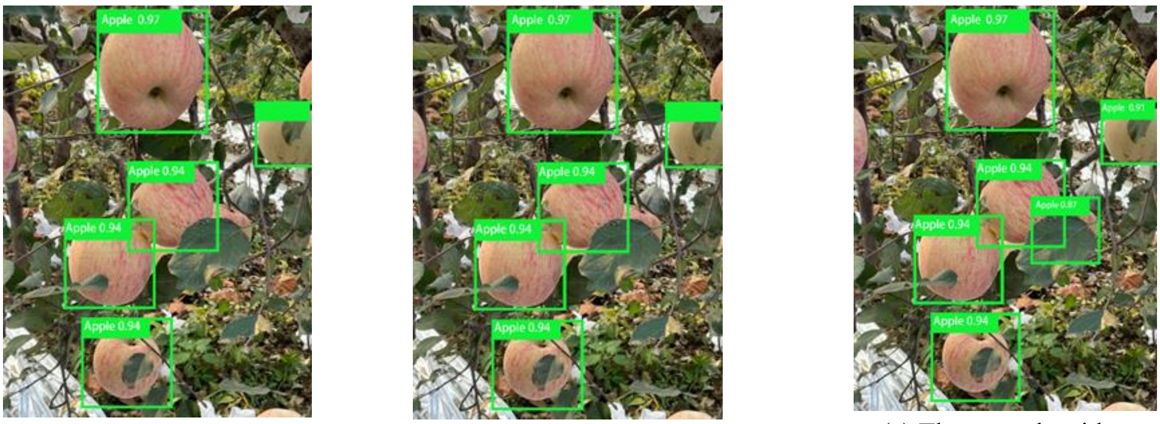
(a) YOLOv3 (b) YOLOv5s (c) The new algorithm

(1) Recognition under sunny and smooth light conditions



(a) YOLOv3 (b) YOLOv5s (c) The new algorithm

(2) Recognition under cloudy conditions



(a) YOLOv3 (b) YOLOv5s (c) The new algorithm

(3) Recognition in leaf-shaded conditions

**Fig. 10 - Apple recognition in different conditions**

The recognition results of the improved YOLOv5 network, YOLOv5s, and YOLOv3 proposed in the study are accurate under sunny skies and smooth light conditions, with no missed recognition. It can be seen that under cloudy conditions, wrong recognition occurs in YOLOv5s and YOLOv3 networks. Under insufficient light, YOLOv3 does not recognize overlapping fruits and smaller fruits in the photos. YOLOv5s only recognizes overlapping fruits but does not recognize smaller fruits. At the same time, YOLOv5s has multiple detections. In the case of leaf occlusion, error recognition occurs in YOLOv5s and YOLOv3 networks. In the case that most of the apple fruits are occluded by leaves, existing algorithms make it difficult to identify apples.

In conclusion, the model obtained after algorithm optimization in this paper can achieve the best detection effect in apple detection, and can effectively prevent false detection and missing detection under the premise of ensuring the recognition accuracy.

**Comparison of different detection algorithms**

To further analyze the recognition performance of the proposed algorithm for Apple targets, this study compares the improved YOLOv5 network with the original YOLOv5s, YOLOv3, and YOLOv4-tiny network models on 200 test set images. The model's mAP value and average recognition speed were used as evaluation metrics. The recognition results, size, and number of parameters for each network model are shown in Table 2.

**Table 2**

**Comparison of test indicators of different models**

| Model | Accuracy / % | recall rate / % | mAP / % | Average Detection Speed (s/pic) |
|---|---|---|---|---|
| The algorithms in this paper | 94.37 | 96.21 | 96.73 | 0.028 |
| YOLOv5s | 90.67 | 93.59 | 94.32 | 0.033 |
| YOLOv4—Tiny | 89.64 | 91.58 | 90.28 | 0.034 |
| YOLOv3 | 87.06 | 92.11 | 87.69 | 0.039 |

As can be seen from Table 2, the improved YOLOv5 network model proposed in this paper has a detection accuracy of 94.37%, a recall of 96.21%, a mAP of 96.73%, and an average single-image detection speed of 0.028 s. Compared with the YOLOv5s, YOLOv4 - Tiny, and YOLOv3 models, the accuracy has been improved by 3.7%, 4.7%, 7.3%, the recall rate increased by 2.6%, 4.6%, and 4.1%, the mAP increased by 2.4%, 6.5% and 9.1%, and the average detection speed of a single image was shortened by 15.2%, 17.6% and 28.2%, respectively. Although the recognition accuracy of the model is slightly affected by the light conditions, branch occlusion, and fruit overlapping, on the whole, compared with other network models, the improved YOLOv5 can effectively recognize the target with strong anti-interference ability, high detection accuracy, and fast recognition speed. Overall, the model proposed in the study is the lightest among the five network models with the highest mAP value. The recognition speed of this model is faster than YOLOv3, YOLOv4, and YOLOv5s networks and can satisfy the requirement of real-time Apple recognition.

## CONCLUSIONS

This study aimed at the detection task of small targets and easily occluded overlapping objects, and improved the traditional YOLOv5 model from three aspects, respectively, by adding the Transformer with attention to the backbone network of YOLOv5, which could form a complementary module with convolution, and the attention mechanism could help the model to better detect apples and such small targets; the neck structure could effectively alleviate the problem of overlap between small targets and targets; the shallow feature P2 with smaller down sampling was added to the head structure for feature output, which was therefore helpful for the positional detection of small target objects. The experimental results showed that the advantages of the apple detection algorithm proposed in this paper were shown in the following two points: first, the detection performance, especially the detection speed of the improved YOLOv5s model was excellent, which was suitable for real-time apple recognition of picking robots; Secondly, the recognition success rate of obscured fruits was also relatively high, 3.7% higher than the original model, and there would be no omission or incorrect recognition.

On the other hand, the apple picking robot can work at night, while the algorithm proposed in the study is designed for fruit recognition in the daytime. Therefore, it may not be suitable for apple target recognition at night, which is the limitation of our detection algorithm.

## REFERENCES


[1] Afzaal H., Farooque A.A., Schumann A.W., Hussain N., McKenzie-Gopsill A., Esau T., Abbas F., Acharya B. (2021). Detection of a potato disease (early blight) using artificial intelligence. *Remote Sens*., 13, 411.

[2] Bresilla K, Perulli G.D., Boini A., et al. (2019). Single-shot convolution neural networks for real-time fruit detection within the tree. *Frontiers in Plant Science*, Vol. 10, 1-12.

[3] Fan Cheng, LI Jianian, et al. (2020). Identification method of strawberry based on convolutional neural network (基于卷积神经网络的草莓识别方法). *Transaction of the Chinese Society for Agricultural Machinery*, Vol. 51, No. 2, 237-244.

[4] Gao F., Fu L., Zhang X. et al. (2020). Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. *Computers and Electronics in Agriculture*, Vol. 176, 1-10.

[5] Gene-Mola J., Sanz-Cortiella R., Rosell-Polo J.R., et al. (2020). Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Computers and Electronics in Agriculture*, Vol. 169, 1-11.

[6] Gene-Mola J., Vilaplana V., Rosell-Polo J.R., et al. (2019). Multi-modal deep learning for Fuji Apple detection using RGB D cameras and their radiometric capabilities. *Computers and Electronics in Agriculture*, Vol. 162, 689-698.

[7] Ghosh, I., Datta Chaudhuri, T. FEB-Stacking and FEB-DNN models for stock trend prediction: A performance analysis for pre and post Covid-19 periods. Decis. Mak. Appl. Manag. Eng. 2020, 4.

[8] Kang H., Chen C., (2019). Fruit detection and segmentation for apple harvesting using visual sensor in orchards. *Sensors*, Vol. 19, No. 20, 45-99.

[9] Kang H., Chen C., (2020). Fast implementation of real-time fruit detection in apple orchards using deep learning. *Computers and Electronics in Agriculture*, Vol. 168, 1-10.

[10] Kang H., Chen C., (2020). Fruit detection, segmentation, and 3D visualisation of environments in apple orchards. *Computers and Electronics in Agriculture*, Vol. 171, 1-10.

[11] Kerkech, M., Hafiane, A.; Canals, R. VddNet: Vine disease detection network based on multispectral images and depth map. Remote Sens. 2020, 12, 3305.

[12] Liu Mochen, Gao Tiantian, Ma Zongxu, et al. (2021). Target detection model of corn weeds in field environment based on MSRCR algorithm and YOLOv4 tiny (基于 MSRCR YOLOv4 tiny 的田间环境玉米杂草检测模型). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 53, No. 2, 246-255, 335.

[13] Long Yan, Li Nannan, Gao Yan, et al. (2021). Apple fruit detection under natural condition using improved FCOS network (基于改进 FCOS 网络的自然环境下苹果检测). *Transactions of the CSAE*, Vol. 37, No. 12, 307-313.

[14] Malinda, M., Chen, J. The forecasting of consumer exchange-traded funds (ETFs) via grey relational analysis (GRA) and artificial neural network (ANN). Empir. Econ. 2020, 3.

[15] Mirko, S., Aleksandar, S., Đorde, S. ANFIS model for the prediction of generated electricity of photovoltaic modules. Decis. Mak. Appl. Manag. Eng. 2019, 2.

[16] Precup, R., Preitl, S.; Petriu, E.; Bojan-Dragos, C.; Szedlak-Stinean, A.; Roman, R.; Hedrea, E. Model-Based fuzzy control results for networked control systems. Rep. Mech. Eng. 2020, 1.

[17] Wang Dandan, He Dongjian, (2019). Recognition of apple targets before fruits thinning by the robot based on R-FCN deep convolution neural network (基于 R-FCN 深度卷积神经网络的机器人疏果前苹果目标的识别). *Transactions of the CSAE*, Vol. 35, No. 3, 156-163.

[18] Wu Xing, Qi Zeyu, Wang Longjun, et al. (2020). Apple detection method based on light-YOLO v3 convolutional neural network (基于轻量化 YOLO v3 卷积神经网络的苹果检测方法). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 51, No. 8, 17-25.

[19] Xia Ye, Xiaohui Lei, Andreas Herbst, Xiaolan Lyu, (2023). Research On Pear Inflorescence Recognition Based on Fusion Attention Mechanism With Yolov5. *INMATEH-Agricultural Engineering*, Vol. 69, No. 1, 11-20. https://doi.org/10.35633/inmateh-69-01

[20] Yan B., Fan P., Lei X. Y., et al. (2021). A real-time Apple targets detection method for picking robots based on improved YOLOv5. *Remote Sensing*, Vol. 13 No. 9, 16-19.

[21] Yan Bin, Fan Pan, Wang Meirong, et al. (2022). Real-time apple-picking pattern recognition for picking robot based onimprovedYOLOv5m (基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别). *Transaction of the Chinese Society for Agricultural Machinery*, Vol. 53, No. 9, 28-38.

[22] Zhang Lu, Huang Lin, Li Beibei, et al. (2021). Fish school counting method based on multi-scale fusion and no anchor YOLO v3 (基于多尺度融合与无锚点 YOLO v3 的鱼群计数方法). *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 52, 237-244.
[23] Zhang Qinghui, Kong Dexiao, Liu Junqiu, et al. (2021). Design of fruit picking system based on inverse kinematics dimension reduction and YOLO v4 (基于逆运动学降维求解与 YOLO v4 的果实采摘系统研究). *Transaction of the Chinese Society for Agricultural Machinery*, Vol. 52k, No. 9, 15-23.
[24] Zhao Hui, Qiao Yanjun, Wang Hongjun, et al. (2021). Apple fruit recognition in complex orchard environment based on improved YOLO v3 （基于改进 YOLO v3 的果园复杂环境下苹果果实识别）. *Transactions of the CSAE*, Vol. 37, No. 16， 127-135.
[25] Zhao Y., Gong L., Zhou B., et al. (2016). Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and color analysis, *Biosystems Engineering*, Vol. 148, 127-137.