A REAL-TIME DETECTION MODEL FOR IDETIFICATION OF CITRUS DURING DIFFERENT GROWTH STAGES IN ORCHARDS
/
一种实时检测模型在果园不同生长阶段的柑橘识别

Changgao XIA *1), Wanlei NI 1), Kun JIANG 1), Xiaofan LI 2) [1]
1) School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China
2) School of Management, Jiangsu University, Zhenjiang 212013, China
Tel: 13815171208; E-mail: xiacg@ujs.edu.cn
Corresponding author: Changgao Xia
DOI: https://doi.org/10.35633/inmateh-68-37

Keywords: *Full growth cycle citrus, Deep learning, Detection, YOLOV-V5*

ABSTRACT
In order to solve the problem of citrus full growth cycle identification in complex scenes, this paper proposed a multi-scale detection model of citrus whole growth cycle in orchard environment. The weighted bi-directional feature pyramid network (BiFPN) is used to combine multiple feature information of high resolution and low-resolution feature layers, and the feature information is extracted by the depth-separable convolution and lightweight New-C3 module. The results show that the average accuracy of the multi-scale detection model proposed in this paper was 91.35%, 92.89%, 94.12%, 90.39% in the young citrus, expanding citrus, ripe citrus and full growth cycle citrus, and the average detection time was 92.60 FPS/s under 1920×1080 image pixels, which meets the real-time detection requirements of citrus orchard.

摘要
为了解决复杂场景下柑橘的全生长周期识别问题，本文提出了一种果园环境下柑橘全生长周期的多尺度检测模型。采用加权双向特征金字塔网络（BiFPN）来融合高分辨率和低分辨率特征层的多项特征信息，且通过深度可分离卷积和轻量型 New-C3 模块实现特征信息的提取。结果表明，本文提出的多尺度检测模型在生长期、膨果期、成熟期和全生长周期柑橘的平均精度为 91.35%，92.89%，94.12%，90.39%，在 1920×1080 图像像素下的平均检测时间为 92.60 FPS/s，满足果园柑橘的实时检测要求。

INTRODUCTION
In recent years, the production of citrus and other fruits has steadily ranked first in the world. As a result, China has gradually become one of the important fruit-producing countries in the world *(Zhao et al., 2016).* Throughout the citrus planting and management process, fruit farmers are the main labour force in orchards, and labour costs account for 50% to 70% of the total input costs. Real-time detection technology can accurately identify citrus at different growth stages, which is one of the important means to realize the fine management of orchards *(Wang et al., 2013; Zhang et al., 2002).*

In the recognition of citrus at different growth stages, there are complex scene problems such as citrus overlap, branches and leaves shading, site influence, light conditions, and the colour and volume of citrus can change with its different growth stages, which can greatly reduce the accuracy of citrus recognition. In the early days, researchers at home and abroad explored many target detection algorithms, and these recognition methods were used to accomplish recognition detection by extracting feature information such as the colour and shape of fruits *(Yu et al., 2013; Kurtulmus et al 2011; Zhao et al., 2016)*. However, these methods can only detect fruits at specific growth stages, and using the same model to detect fruits at different growth stages is not satisfactory *(Illingworth et al., 1988)*. *Rakun et al., (2011),* used colour segmentation to separate the fruit from the background in order to overcome the influence of light conditions and partial fruit occlusion. By analysing three different features (colour, texture and 3D space) of possible regions, the detection rate of fruits was improved. However, the detection effectiveness of this method is greatly reduced when there are more small targets in the image. The recognition effect of the above method is highly dependent on the environment and influenced by the lighting conditions. When the detected fruits are small, dense, or the colour features are similar to the background, it is difficult to extract the feature information, resulting in lower detection accuracy.

[1] *Changgao Xia, Prof. Ph.D. Eng.; Wanlei Ni, M.S. Stud. Eng.; Kun Jiang, M.S. Stud. Eng.; Xiaofan Li, M.S. Stud Eng.*

Since the 1980s, machine learning has been widely applied in the direction of agriculture *(Tanima et al., 2020).* Many machine learning theories have been applied to fruit recognition in orchards, such as Canny, HOG (Histogram of Oriented Gradient), and SVM (Support Vector Machine) methods to extract fruit features and train the features, and then use the trained model to predict the location and class of new input features. *Kelman et al., (2014),* used the Canny edge detection operator to extract all the contour information in the image and construct the 3D contour feature function of Apple, and then used this function to screen all the contour information in the image to identify apple. After experimental analysis, this function can identify 94% of the apples. But a branch or leaf with a convex surface similar to the contour of an apple resulted in 14 percent of false positives. *Ji et al., (2012),* used SVM+HOG classifier to classify bagged apples and ordinary apples, achieving 89% detection accuracy for bagged apples. However, the average detection time of this algorithm is 352ms, which makes it impossible to implement this algorithm in real-time system, and the average error rate is 11% due to leaf occlusion. *Zhuang et al., (2018),* proposed a robust detection algorithm based on monocular vision system to identify citrus fruits. Combined with Otsu threshold, morphological operation, watershed transformation of marker control and other methods, the potential citrus region was located from the colour map, and then the local texture information was extracted from the potential region. Finally, SVM was used to make the final decision. The results showed that the recall rate of citrus was greater than 0.86, but the detection time was 0.53s. In summary, the method of machine learning for fruit recognition has low accuracy and is easily disturbed by the complex environment of the orchard.

In recent years, convolutional neural network has been rapidly developed in the field of target recognition *(Liu et al.,2020; Zheng et al., 2021)*, and more and more researchers worldwide have applied those recognition algorithms to orchard crop recognition and achieved remarkable results.

Tian et al., (2019), proposed a method using DenseNet to improve YOLO-V3 model for detecting apples during different growth stages. The DenseNet method is used to process the lower resolution feature layers in the YOLO-V3 network, which can effectively enhance feature propagation, promote feature reuse and improve network performance. Experiments showed that the F1 scores of apples in the young apple, expanding apple, ripe apples and full growth apple were 0.832, 0.841, 0.864, and 0.817. Respectively, the average detection time under the image with a resolution of 3000×3000 was 0.304s. This method can complete the study of the entire growth cycle of apples, but the recognition accuracy and average detection time still need to be further improved. *Wang et al. (2021)* used the pruned YOLO-V5s model to identify apples. The experiments showed that the apple recall, accuracy, F1 score and false detection rate of small fruits reached 87.6%, 95.8%, 91.5%, and 4.2%. The average detection time per image is 8ms, and the model size is only 1.4MB, but the model is not good for the recognition of occluded apple fruit. *Lyu et al., (2022),* proposed a lightweight YOLOv5-CS model to identify and count green citrus in natural environment. The model uses the CBAM attention mechanism to optimize the convolutional and detection layers, while using CIoU (Complete Intersection over Union) loss function and cosine annealing algorithm to strengthen the model, and finally implant the model into a mobile device. The results show that the mAP@0.5 of the YOLOv5-CS model for green citrus is 98.23%, the recall rate is 97.66%, the inference speed of detecting pictures on the server is 0.017s, and the inference speed on Nvidia Jetson Xavier NX is 0.037s. In the follow-up research, the citrus at different growth stages can be studied. *Huang et al., (2021),* improved the YOLO-V5s model using the CBAM attention mechanism and adaptively fused features, and then pruned the model to reduce the model size. The experiment was carried out on a self-made citrus dataset, the detection accuracy was 93.32%, and the processing speed of the edge computing device was 180ms/frame.

To sum up, those neural network recognition model has the advantages of high detection accuracy and fast detection speed in the detection of orchard crops. However, the above studies are not effective in detecting multi-size orchard crops and dense scenes. Meanwhile, the full growth cycle of citrus fruits has not been studied. Therefore, this paper constructs a multi-scale detection model for the whole growth cycle of citrus in the orchard environment, and explores the recognition problems in complex scenes such as citrus with different sizes, different colours, overlapping citrus, and occlusion of branches and leaves. In this paper, the following two structural optimizations are carried out for the YOLO-V5m model:

(1) Optimization of the YOLO-V5m neck network using a weighted bi-directional feature pyramid network (BiFPN).

(2) Deeply separable convolution and lightweight New-C3 modules are proposed to replace the conventional convolution and C3 modules for feature information extraction.

The research results show that the YOLOV5_L model has the advantages of high detection accuracy, fast detection speed, and small model size. Firstly, the adopted weighted bidirectional feature pyramid network fuses multiple feature information from high-resolution and low-resolution feature layers to enhance the recognition accuracy of the model for citrus. Secondly, the lightweight module used optimizes the YOLO-V5 network structure, which greatly reduces the model size.

MATERIALS AND METHODS

Image data preprocessing

In this study, the collection site of citrus images was located in Jiangxinzhou citrus picking garden, Zhenjiang City, Jiangsu Province. The collected images contain orchard scenes such as unobstructed, branches occlusion, leaves occlusion, branches, and leaves occlusion and citrus overlapping. In order to identify full growth cycle citrus, there are 500 original images were collected for each growth stage, and the three growth stages formed a total original dataset of 1500 images.

In this paper, data augmentation methods such as Mosaic, Augment HSV, and Random affine were used to increase the richness of the dataset. First, the Mosaic data enhancement method is used to randomly select four images and randomly scale them, and then randomly stitch them together, which greatly increases the richness of the dataset. Second, the Augment HSV method is used to randomly adjust the chroma, saturation, and lightness of the image to increase the diversity of the image. Finally, random affine swaps are performed via the Random affine method. In this paper, the above image enhancement methods greatly enriched the types and sizes of images. The original 500 pictures in each growth stage were expanded to 1500 pictures, and the total datasets reached 4500 pictures. After that, 150 images are randomly taken as the test dataset in each growth stage, and the remaining 4050 images are used as the training set to train the model. The rich dataset makes the network more robust.

Basic model selection

The YOLO-V5 network improves the network structure and training methods on the basis of the YOLO-V4 network and improves the detection performance of the model. In the model image processing method, as show in Fig.1, YOLO-V5 uses the adaptive image scaling function to process the original images of different sizes and obtain various scaling coefficients, then selects the smallest coefficient to multiply with the length and width of the original image, and finally adaptively adds the least black edge to the original image to reduce the redundant information, which reduces the computation and improves the detection speed in the inference process.

YOLO-V5m is a lightweight detection model designed for mobile devices, which is difficult to accurately detect small targets and perform deeper feature fusion in complex orchard environments. The young citrus contains a large number of small citrus, and as the tool moves during the shooting of citrus, the citrus in the distance will become smaller and there will be pixel distortion, which makes the fusion of feature information very difficult. To address these issues, this paper optimizes the YOLO-V5m model to strengthen the detection of small targets and the fusion of multi-scale feature maps.

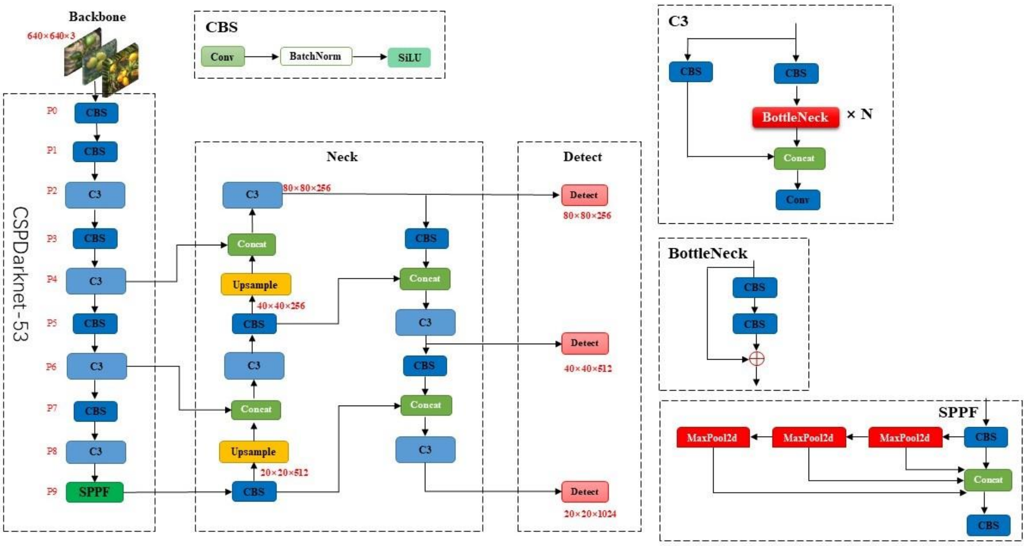

Fig. 1 - Structure of Yolo-V5 model

Model design of weighted bi-directional feature pyramid network (BiFPN)

The orchard environment is very complex, the citrus is small and there is pixel distortion in the distant citrus as the shooting tool moves. the YOLO-V5m model cannot extract the feature information of this part of the citrus, which leads to many missed and wrong detections. To this end, this paper proposes an optimized YOLO-V5m model based on a weighted bi-directional feature pyramid network *(Tan et al., 2020)* (BiFPN) structure, which enhances the fusion of feature information. As shown in Fig.2, the main idea of the BiFPN network structure is divided into two parts: one is efficient bidirectional cross-scale connectivity, efficient bidirectional cross-scale connection, and the other is the fusion of weighted feature maps. For bidirectional cross-scale connectivity, this network structure is equipped with bottom-up and top-down bidirectional channels. Different scale feature layers use upsample and convolution with a stride of 2 to adjust the resolution, so as to achieve high resolution. The fusion of high-resolution and low-resolution feature layers, and adding lateral connections between the original input and output nodes of the same feature, fuses more clementine feature information at a very low computational cost.

Since the input feature layers of different scales have different resolutions and contain different amounts of feature information, the input feature layers of different scales contribute unevenly to the output feature layers. In order to solve this problem, the BiFPN network structure needs to weigh the different scale feature layers during the feature layer fusion. Equation (1) is a fast normalized fusion method.

The ReLU activation function ensures that a small value is used to ensure the stability of the value. In order to further improve the efficiency, the Depthwise Separable Convolution is used here for feature fusion, and the batch normalization layer and the ReLU activation layer are added after each convolution operation.

$$O = \sum_i \frac{\omega_i}{\varepsilon + \sum_j \omega_j} I_i \tag{1}$$

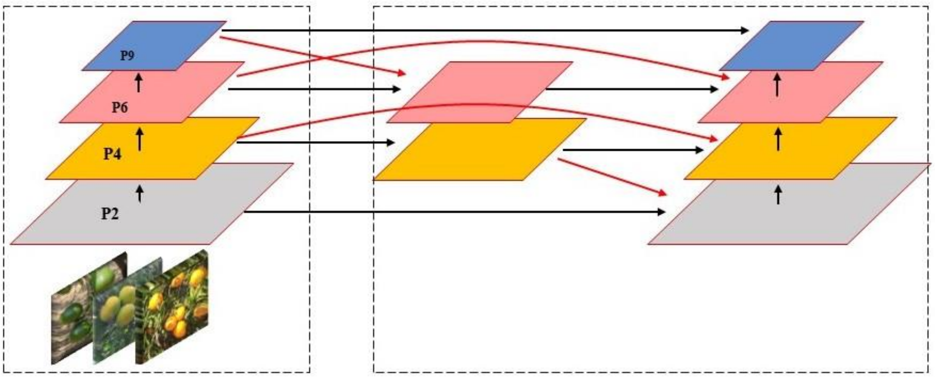

Fig. 2 - BiFPN feature network

Model design of light

Depthwise Separable Convolution

The YOLO-V5m model uses Standard Convolution (CBS) for the extraction of feature information, which causes the network to generate a large number of parameters. The characteristics of citrus fruits are relatively single and do not require too many parameters to fit the data. To this end, this paper proposes a feature extraction module based on Depthwise Separable Convolution (DWC) to optimize the YOLO-V5m model. Depthwise Separable Convolution is a lightweight network structure, which has the advantages of less number of parameters and lower operational cost compared with standard convolution. As shown in Fig.3, the Depth-Separable Convolution consists of two parts: Depthwise Convolution and Pointwise Convolution. One convolution kernel is responsible for one channel, and the number of feature map channels generated during the convolution process is exactly the same as the number of input channels. The point-by-point convolution is actually a 1×1 convolution, which plays two roles in the depth-separable convolution: first, it allows the depth-separable convolution to freely change the number of output channels; the other is to output the channel-by-channel convolution. feature layer for channel fusion.

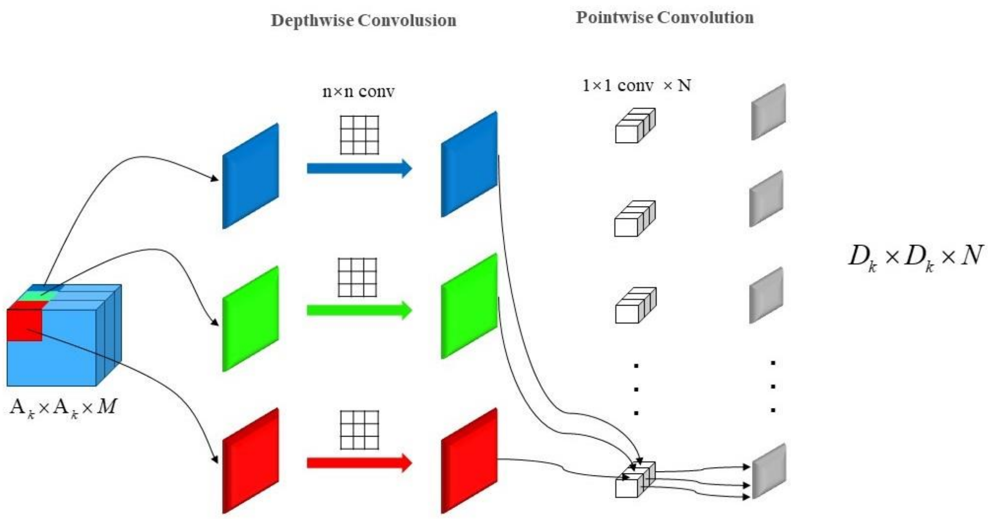

Fig. 3 - Depthwise Separable Convolution model

In terms of computational performance, the Depthwise Separable Convolution has been greatly improved compared to the standard convolution. Assuming that, the size of the input feature map is $A_k \times A_k \times N$, the size of the output feature map is $D_k \times D_k \times M$, and the size of the convolution kernel is *nxn*. The standard convolution layer operates the input feature map of each channel using the convolution kernel before outputting, and the computational volume is as follows:

$$C_{std} = D_k \times D_k \times M \times N \times n \times n \tag{2}$$

Depthwise Separable Convolution is first performed channel-by-channel and then point-by-point convolution, which is computed as:

$$N_{std} = D_k \times D_k \times M \times n \times n + D_k \times D_k \times M \times N \tag{3}$$

from the above equation, the ratio of the DWC to the standard convolution as:

$$\frac{N_{std}}{C_{std}} = \frac{D_k \times D_k \times M \times n \times n + D_k \times D_k \times M \times N}{D_k \times D_k \times M \times N \times n \times n} = \frac{1}{N} + \frac{1}{n^2} \tag{4}$$

In general, both N and n are greater than 1. From equation (4), the computational efficiency of the Depth Separable Convolution is better than that of the standard convolution.

New-C3 model

As shown in Fig.1, the C3 module of the YOLO-V5m model is the main module for learning the residual features. The structure of the C3 module is divided into two branches, one uses multiple Bottleneck stacks and 3 standard convolutional layers, the other passes through a basic convolution module, and finally, the two branches are concated. The standard convolution process is "convolution" - "batch normalization" - "non-linear activation", which generates a large number of parameters and increases the computational effort. In order to streamline the network structure, reduce the computation and decrease the model inference time, this paper proposes a lightweight New-C3 module for learning the residual features.

Compared with C3, the biggest change of the New-C3 module is to use NewBottleNeck to replace the original BottleNeck. As shown in Fig.4, the New-C3 module is a phased convolution module that uses Depthwise Separable Convolution and NewConv convolution instead of standard convolution to obtain more feature maps and eliminates redundant features. The function of NewConv is to reduce the number of channels with similar feature maps during the convolution process. For these repeated channels, using standard convolution calculation will consume a lot of computing power. NewConv is also divided into two branches, one uses the standard convolution unit to obtain the first part of the channel, the other one uses the Depthwise Separable Convolution on the first part of the channel to obtain the remaining part of the channel, and the other uses the Depthwise Separable Convolution to obtain all channels. Finally, the two parts of the channel are spliced, which reduces the amount of calculation and avoids the appearance of duplicate channels.

The NewBottleNeck module structure is also divided into two branches and is greatly affected by the Stride. For the case of Stride=1, in one branch, a NewConv is used as an extension layer to increase the number of channels, and then the NewConv module is used to reduce the number of channels to match the number of channels in the Shortcut layer. In the other branch, Depthwise Separable Convolution is used for feature extraction, and the number of channels is matched. Finally, the outputs of the two branches are connected using Shortcut. For Stride=2, between the two NewConv of the first branch, a Depthwise Separable Convolution with Stride=2 is used for connection, and Depthwise Separable Convolution and standard convolution unit are used on the other side for feature extraction. Finally, the outputs of the two are connected by shortcut. This model structure can obtain more feature maps while reducing the size and number of parameters of the model.

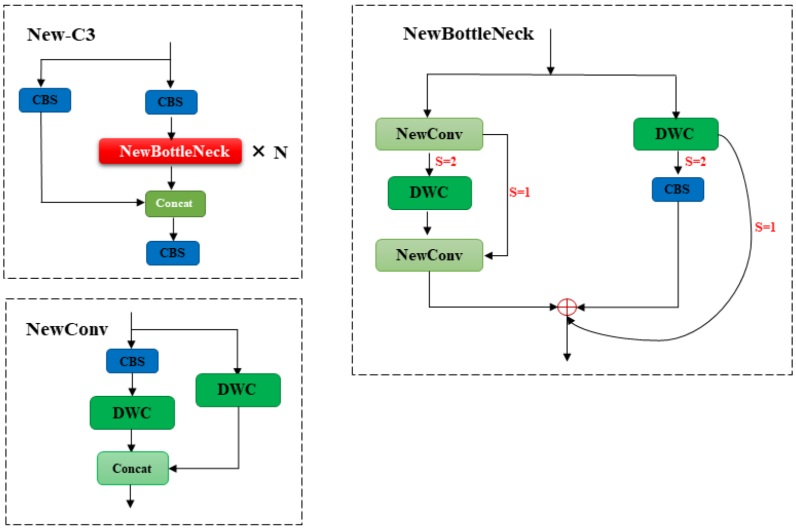

Fig. 4 - New-C3 model

RESULTS

Evaluation indicators

The precision P, recall R, average precision (AP), F1 Score are selected in the field of target detection. Accuracy indicates the probability of successful prediction among the predicted targets. Recall indicates the probability of successful prediction among the targets to be predicted. The average accuracy AP represents the cumulative sum of accuracy and recall under different confidence thresholds. F1 is a composite performance indicator that reconciles accuracy and recall, and the more F1 converges to 1, the better the overall performance of the model. In this paper, the threshold value is set to 0.5, and the prediction is considered successful if the intersection ratio between the measured target and the real target is greater than 0.5.

Precision:

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

Recall:

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

Average Precision:

$$AP=\sum_{K=i}^{N}P(k)\Delta R(k) \tag{7}$$

F1:

$$F_1=\frac{2PR}{P+R} \tag{8}$$

Ablation Experiment

In this paper, based on the YOLO-V5m model, each optimization part was compared separately, and a series of ablation experiments were performed on the data sets of various growth stages of citrus to test the performance of the model. The specific experimental results are shown in Table 1.

The following results can be observed from the data in Table 1:

(1) Full growth cycle citrus

The average accuracy of YOLO-V5m in full growth cycle citrus was 87.54%. When using only BiFPN optimization, the average accuracy increases by 3.14%, the number of parameters increases by 2.29M, and the detection speed decreases by 13.2PFS/s. In order to reduce the number of parameters and speed up the detection speed of the model, this paper improves the lightweight network structure for feature extraction and fusion. The parameter amount of the model is reduced to 9.23M, and the detection speed is increased to 92.6PFS/s, which greatly improves the detection speed, and the AP is improved the average accuracy by 2.85%.

(2) Young citrus

The average precision of YOLO-V5m was 88.64% for citrus in the young citrus. When only BiFPN was added, the model accuracy increased by 3.08%. After combining the BiFPN and DG model, the most significant improvement in average accuracy was increased by 2.71%.

(3) Expanding citrus

The average accuracy of YOLOv5m in expanding citrus was 90.23%. When only using BiFPN optimization, the average accuracy increased by 2.79%. After combining the BiFPN and DG model, the most significant improvement in average accuracy was increased by 2.66%.

(4) Ripe citrus

The average precision of YOLO-V5m in ripe citrus was 91.24%. When only using BiFPN optimization, the average accuracy increased by 3.34%. After combining the BiFPN and DG model, the most significant improvement in average accuracy was increased by 2.88%.

It was found that the following two improvements were made in this paper relative to the baseline YOLO-V5m model. Firstly, after adding the optimization of the network using BiFPN showed that the AP values of expanding citrus, ripe citrus, and full growth cycle citrus increased significantly. BiFPN enables the model to efficiently realize the bidirectional cross-scale connection of feature layers and weighted feature fusion, as also a weighted fusion of feature information of different resolutions. The large size and fruit color of expanding citrus differ slightly from the surrounding environment. At the ripe citrus, the citrus volume is large, and the fruit color is significantly different from the surrounding environment. Therefore, the BiFPN structure can obtain more volume and color features, so that the model can better recognize the citrus in the expanding citrus and ripe citrus. When using both the BiFPN and DG model to optimize the YOLO-V5m model, the model has achieved a good detection for citrus at all growth stages, but both improvements will make the model parameter quantity increase and reduce the model detection speed.

Table 1

Model Ablation Experiment Analysis Results

Citrus Stage	BiFPN	DG	AP / %	FPS / s	Parameters (M)
Full growth cycle citrus			87.54	71.40	19.91
	√		90.68	58.20	22.20
	√	√	90.39	92.60	10.68
Young citrus			88.64	71.40	19.91
	√		91.72	58.20	22.20
	√	√	91.35	92.60	10.68
Expanding citrus			90.23	71.40	19.91
	√		93.02	58.20	22.20
	√	√	92.89	92.60	10.68
Ripe citrus			91.24	71.40	19.91
	√		94.58	58.20	22.20
	√	√	94.12	92.60	10.68

Influence of data category

In order to verify the influence of data categories on the detection performance of the model, the YOLOV5_L model was used in this paper to analyse the citrus data sets of the young citrus, expanding citrus, ripe citrus, and full growth cycle citrus. The obtained F1 score is shown in Table 2, and the P-R curve is shown in Fig.5. The P-R curve is an important index to evaluate the detection results of the identification model, and the area enclosed with the coordinate axis is the AP value. It is generally believed that the closer the P-R curve of the model is to the upper right corner of the coordinate axis, the better the performance of the model. The F1 score of the full growth cycle citrus data set was 0.845, which was lower than 0.017 in the young citrus, 0.063 in the expanding citrus, and 0.067 in the ripe citrus, and the P-R curve is at the lower left of each growth stage, indicating the detection performance of this model for the full growth cycle of citrus significantly lower than that of citrus at different growth stages. This is because of the wide variety of data for full growth cycle citrus and the inconsistency of citrus growth characteristics at each growth stage, so the detection difficulty is greatly increased during the data enhancement process of the model and the training process of the model. Among the individual growth stages, due to the small size of the young citrus and the similar colour to the surrounding environment, the F1 scores in the young citrus were significantly smaller than those in the expanding and ripe citrus, and the P-R curve was at the lower left of the expanding and ripe citrus. The F1 score and the P-R curve of the fruit in the expanding citrus are slightly lower than those of the fruit in the ripe citrus because the size of the fruit in the expanding citrus is the same as that in the ripe citrus, but the colour is between green and yellow, which is slightly similar to the surrounding branches and leaves. YOLOV5_L model has the best recognition effect on ripe fruits, which is due to the large size of ripe fruits and the colour characteristics that are different from the colour of the surrounding branches and leaves, so this model has the best effect on the detection of ripe citrus.

Table 2

F1 Scores of Citrus Detection Models in Several Categories

Class	F1
Young citrus	0.862
Expanding citrus	0.908
Ripe citrus	0.912
Full growth cycle citrus	0.845

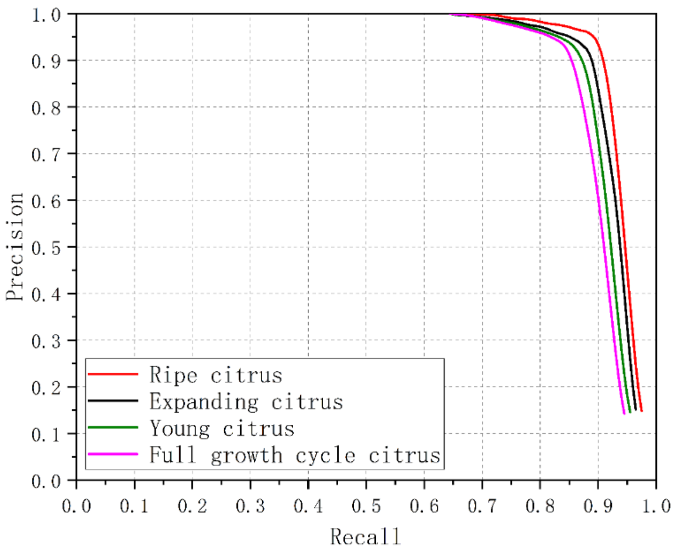

Fig. 5 - P-R Curves of Citrus Detection Models in Several Categories

Comparison of different algorithms

In this section, in order to further the superiority of YOLOV5_L model for citrus recognition effect. The YOLOV5_L model is compared with YOLO-V5s with faster detection speed and smaller memory; YOLO-V4, Faster R-CNN model with higher accuracy but slower detection speed.

The comparison results are shown in Table 3. From Table 3, it is found that the YOLOV5_L model has superior performance in detection accuracy with an AP of 90.39%, which is 2.71%, 3.07%, and 4.07% higher compared to Faster R-CNN ResNet50, YOLO-V4, and YOLO-V5s. This is because the full growth cycle citrus dataset contains a large number of small green citrus, large green citrus, and large yellow citrus. The YOLOV5_L model adds an optimized feature fusion network based on BiFPN, so it has more recognition ability for the above mentioned full growth cycle citrus. In terms of detection speed, this paper proposes a lightweight network structure for feature information extraction and fusion, and the experimental results show that the detection speed is improved to 92.60 FPS/s, which is 2.80 times, 2.42 times, and 1.02 times higher compared with Faster R-CNN ResNet50, YOLO-V4, and YOLO-V5s. In terms of model size, the model size of YOLOV5_L model is reduced to 10.68MB, which is much smaller than other models and can be flexibly deployed in low-memory devices.

Table 3

Performance Results from Different Object Detection Algorithms.

Models	AP %	FPS / s	Size / MB
Faster R-CNN ResNet50	87.68	33.12	314.06
YOLO V4	87.32	38.20	236.48
YOLO V5s	86.32	90.91	14.41
YOLOV5_L	90.39	92.60	10.68

CONCLUSIONS

Firstly, this paper uses a weighted bidirectional feature pyramid network structure to fuse feature layers with different resolutions, so that the model learns more feature information during the training process. The experimental results show that the multi-scale detection model of the citrus full growth cycle in an orchard environment constructed in this paper has better detection accuracy compared with previous detection networks. On the one hand, the network has good detection performance in detecting citrus with different degrees of shading and overlapping; on the other hand, the network has a good detection effect even in the growth period when the citrus is small and the colour characteristics of citrus are similar to those of citrus leaves. The future work is to apply the model proposed in this paper to computer devices with mobile platforms and to track and identify the fruits to realize the yield assessment of citrus in orchards and really apply the model.

ACKNOWLEDGEMENT

I mainly completed the design of experiments and the writing of papers, Changgao Xia provided funding, and others provided trial assistance. Project source was the Major Science and Technology Innovation Project of China: Research and development of high-torque reserve diesel engine and multi-point power output transmission for hilly and mountain tractors (1621120014).

REFERENCES

[1] Huang, H., Huang, T., Li, Z., Lyu, S., & Hong, T. (2021). Design of Citrus Fruit Detection System Based on Mobile Platform and Edge Computer Device. *Sensors*, 22(1), 59.

[2] Illingworth, J., & Kittler, J. (1988). A survey of the Hough transform. *Computer vision, graphics, and image processing*, 44(1), 87-116.

[3] Ji, W., Zhao, D., Cheng, F., Xu, B., Zhang, Y., & Wang, J. (2012). Automatic recognition vision system guided for apple harvesting robot. *Computers & Electrical Engineering*, 38(5), 1186-1195.

[4] Kurtulmus, F., Lee, W. S., & Vardar, A. (2011). Green citrus detection using ‘eigenfruit’, color and circular Gabor texture features under natural outdoor conditions. *Computers and Electronics in Agriculture*, 78(2), 140-149.

[5] Kelman, E. E., & Linker, R. (2014). Vision-based localisation of mature apples in tree images using convexity. *Biosystems Engineering*, 118, 174-185.

[6] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International journal of computer vision*, 128(2), 261-318.

[7] Lyu, S., Li, R., Zhao, Y., Li, Z., Fan, R., & Liu, S. (2022). Green Citrus Detection and Counting in Orchards Based on YOLOv5-CS and AI Edge System. *Sensors*, 22(2), 576.

[8] Rakun, J., Stajnko, D., & Zazula, D. (2011). Detecting fruits in natural scenes by using spatial-frequency based texture analysis and multiview geometry. *Computers and Electronics in Agriculture*, 76(1), 80-88.
[9] Tanima, D., Ranjit, K., Bhar, L. M. (2020) Application of Machine Learning Techniques with GARCH Model for Forecasting Volatility in Agricultural Commodity Prices[J]. *Indian Society of Agricultural Statistics*, 74(3), 187-194
[10] Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and electronics in agriculture*, 157, 417-426.
[11] Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781-10790).
[12] Wang, Q., Nuske, S., Bergerman, M., & Singh, S. (2013). Automated crop yield estimation for apple orchards. In *Experimental robotics* (pp. 745-758). Springer, Heidelberg.
[13] Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, 210, 271-281.
[14] Yu, Z., Cao, Z., Wu, X., Bai, X., Qin, Y., Zhuo, W., ... & Xue, H. (2013). Automatic image-based detection technology for two critical growth stages of maize: Emergence and three-leaf stage. *Agricultural and forest meteorology*, 174, 65-84.
[15] Zhao, Y., Gong, L., Huang, Y., & Liu, C. (2016). A review of key techniques of vision-based control for harvesting robot. *Computers and Electronics in Agricultur*e, 127, 311-323.
[16] Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and electronics in agriculture*, 36(2-3), 113-132.
[17] Zhao, C., Lee, W. S., & He, D. (2016). Immature green citrus detection based on colour feature and sum of absolute transformed difference (SATD) using colour images in the citrus grove. *Computers and Electronics in Agriculture*, 124, 243-253.
[18] Zhuang, J. J., Luo, S. M., Hou, C. J., Tang, Y., He, Y., & Xue, X. Y. (2018). Detection of orchard citrus fruits using a monocular machine vision-based method for automatic fruit picking applications. *Computers and Electronics in Agriculture*, 152, 64-73.
[19] Zheng, C., Chen, P., Pang, J., Yang, X., Chen, C., Tu, S., & Xue, Y. (2021). A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard. *Biosystems engineering*, 206, 32-54.