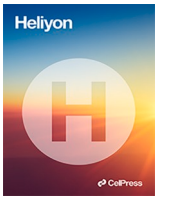
Heliyon 10 (2024) e29035

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Multiscale apple recognition method based on improved CenterNet

Han Zhou

College of Mechanical and Electrical Engineering, Hainan Vocational University of Science and Technology, Haikou, 571126, Hainan, China

ARTICLE INFO

Keywords:
Apple recognition
Improve YoloV5
ROI
Resnet-44

ABSTRACT

Traditional apple-picking robots are unable to detect apples in real-time in complex environments. In order to improve detection efficiency, a fast CenterNet apple recognition method for multiple apple targets in dense scenes is proposed. This method can quickly and accurately identify multiple apple targets in dense scenes. The backbone network mainly consists of resnet-44 fully convolutional network, region of interest network (RPN), and region of interest (ROI). The experimental results show that the improved YoloV5 network model has a higher recognition accuracy of 94.1% and 95.8% for apple in the night environment, which improves the recognition accuracy of the occluded features and the features in the dark light, and the model is more robust in the actual data set.

1. Introduction

At present, agricultural production continues to develop in the direction of scale, intensification and precision. The demand for intelligent and automated agricultural equipment is also increasing rapidly. Apple is the most productive fruit in China. Due to the complex orchard environment, we still rely on human workers to pick it. Therefore, under the circumstances of a shortage of agricultural labor force and increasing picking cost, it is of great practical significance and broad application prospect to replace manual picking with the apple-picking robot. The detection of apple target is the core technology of robot picking. Manual harvesting is not only labor intensive, but also labor shortage with the reduction of the rural labor force. It is urgent to develop apple-picking robots to reduce excessive dependence on labor [1]. Accurate and rapid identification of apple targets are important prerequisite for robots to realize independent picking [2]. In the orchard environment, dense scenes refer to images taken from a long distance that contain a large number of fruits or overlap and occlusion of fruits. The recognition of multiple apple targets in dense scenes are very important to improve the recognition efficiency of the picking robot and realize the intelligent picking of apples [3].

Traditional fruit recognition methods are mainly based on the color, texture and shape characteristics of fruits [4]. This kind of method has a good recognition effect on single fruit or adjacent fruits, but the recognition accuracy of fruits overlapping each other or blocked by branches and leaves in orchards is reduced [5]. In recent years, Convolutional Neural Networks (CNN) has excellent performance in target detection and has been widely used in fruit recognition [6]. In principle, fruit recognition can be divided into two categories. Firstly, the candidate areas that may contain fruits are generated through network, and then the candidate areas are classified. This kind of the network generally takes a long time to recognize [7]. The other methods give confidence and position coordinates of the fruit directly through CNN, which is characterized by improving the speed of fruit identification, but also using an Anchor box to guide the accuracy of fruit identification of occlusion [8]. In the prediction stage, it is necessary to delete the repeated

E-mail address: 121746086@qq.com.

https://doi.org/10.1016/j.heliyon.2024.e29035
Received 27 August 2023; Received in revised form 28 March 2024; Accepted 28 March 2024
Available online 2 April 2024

candidate frames by Non-Maximum Suppression (NMS), so that the target recognition time is longer. However, the complex background environment, make it very difficult to identify small fruits in the early stage of fruits and vegetables. Extensive research has been done on fruit identification at home and abroad [9–12]. The fruit target recognition methods mainly include color difference method [13], K-means clustering method [14], fuzzy C-means method [15], K-Nearest Neighbor method, artificial neural network [16]. Although the above methods can recognize the fruit target in the image, they are all based on the color, shape or texture characteristics of the fruit. For the target fruit with a big color difference from the background, it is a simple method to extract the target fruit area in the image by using color features [17]. Traditional deep learning target detection algorithms are mainly divided into two types: one is the one-stage target detection method, which does not generate candidate regions and has a fast detection speed [18]. The representative algorithms are the You Only Look Once (YOLO), Single Shot MultiBox Detector (SDD) and so on [19]. Tian et al. [20] tested apples based on the YOLOv3. Yue et al. [21] realized the identification of citrus by improving the YOLO network. Kuznetsova et al. [22] combined YOLOv3 with DenseNet to realize the detection of litchi strings. Jing et al. [23] proposed an improved multi-scale IMS-YOLO to detect apples. Mondino et al. [24] used the SSD algorithm to quickly detect the quality of litchi. The other is the two-stage target detection method, which realizes detection by classifying the obtained suggested areas. The detection speed is low, but the accuracy is high. The representative algorithms are Faster R–CNN, Mask R–CNN, Cascade RCNN [25]. Qummar et al. [26] realized the fruit identification of Rosa based on Faster R–CNN. Jing et al. [27] used the improved Cascade RCNN network to distinguish green apples from apples in different the mature stages, but did not further distinguish apples in color-changing stage and the mature stage. The product that reaches a certain proportion in the product detection box is taken as the target output, thus reducing errors [28]. Rehman et al. [29] proposed a parallel framework for real-time identification and classification of apple leaf disease. It also has important value for apple detection.

This paper proposed a multiscale apple recognition method based on improved CenterNet. The detection accuracy is improved by data enhancement, and the improved CenterNet model can realize apple ontology feature recognition in the complex environment at night. Compared with other traditional methods, it has advantages in picking efficiently at night and during the day.

2. Methods

2.1. CenterNet network

CenterNet adopts the idea of "point is target", and determines the target by finding the CenterNet point. CenterNet preprocesses the input image and downsampling by 4 times. The reserved key points are screened by Max pooling, and a prediction box is generated near the key points. Net network identification process is shown in Fig. 1. The total number of output channels of the network is $C+4$, C is the number of categories to identify the target, and 4 is the number of channels, indicating the width and height of the target and the horizontal and vertical coordinate errors of the center point of the target.

2.2. Identification and positioning method of apple string picking point

To solve the problem of identifying and locating the picking points of Apple bunches under complex background, the algorithm flow is shown in Fig. 2. It can be divided into quick identification of the picking fruit stalks through the YOLOv5 target detection algorithm and the connectivity between Apple bunches. According to the coordinates of the picking point, the picking robot is guided to carry out three main parts of the picking operation.

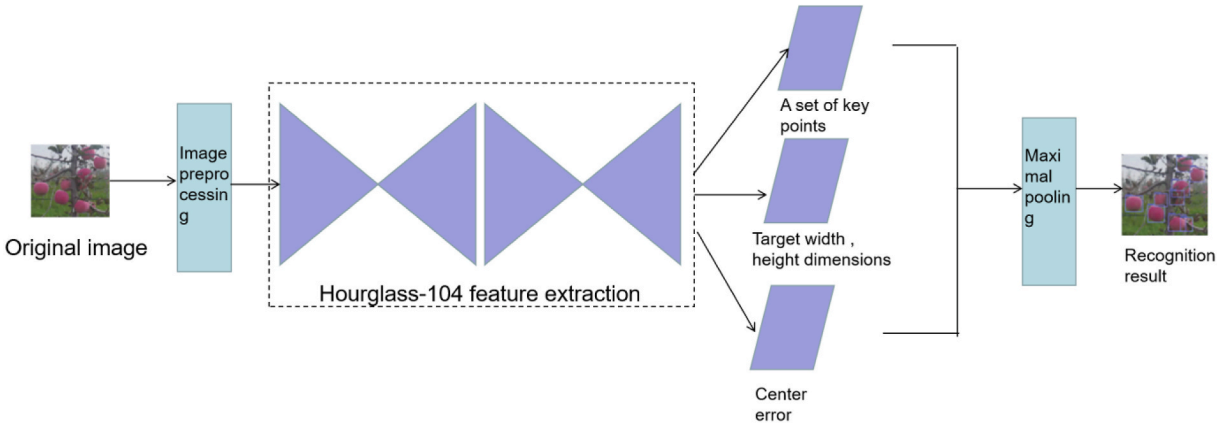


Fig. 1. Frame of CenterNet.

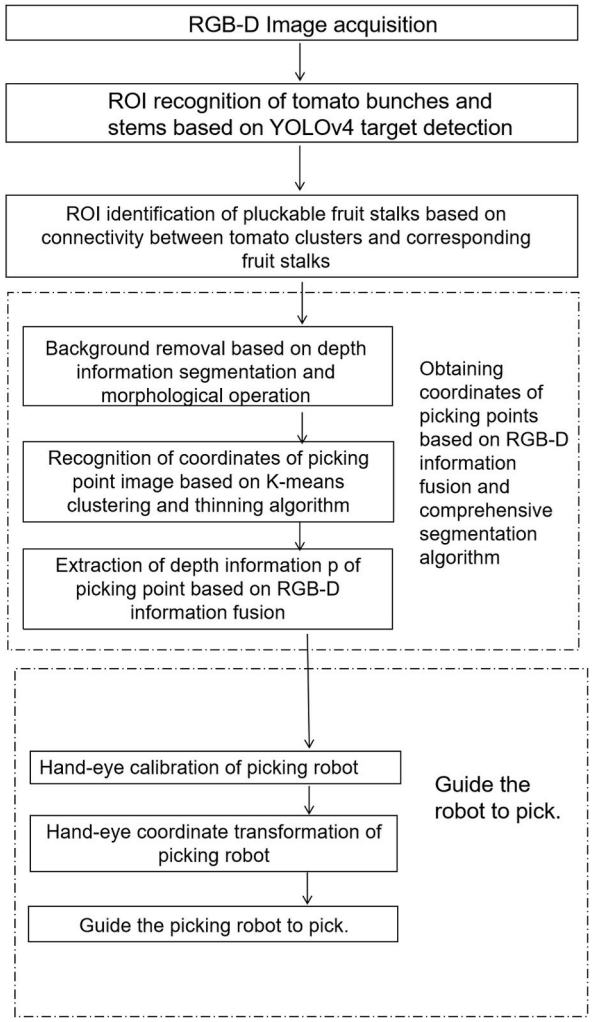


Fig. 2. Technological process for recognizing and locating Apple cluster picking.

2.3. CenterNet improved design of the network

Backbone network is an important part of the convolutional neural network, which mainly extracts the features of targets. The network consists of one or more hourglass modules, each hourglass module extracts features from the input image by down-sampling

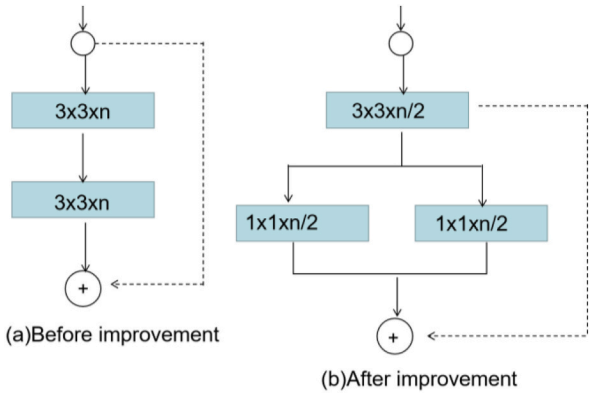


Fig. 3. Residual module improvement.

and up-sampling. Both the up-sampling and down-sampling processes adopt multiple Residual modules, and the network depth is 104 layers. Its structure is complex and huge, and it is restricted by huge parameters. In view of the fact that the recognition in this study is a single category of Apples, and the shallow network can also realize the feature extraction of Apple targets, this paper improves and designs a lightweight CenterNet backbone network to improve the recognition speed of Apples. Considering that Apples are mostly small targets in dense scenes, the information about small target Apples may be lost if the resolution. Only down-sampling the image by 3 times to reduce the resolution of the feature map, and then restores the resolution by 3 times upsampling. In order to reduce the amount of the network parameters and improve the speed of Apple target recognition. The channel number of 3×3 convolution in the first layer of the original residential module is compressed by 2 times, and the 3×3 convolution in the second layer is replaced by 1×1 block convolution. The improved design of the improved Residual module is shown in Fig. 3.

The redesigned hourglass module is named Tiny Hourglass, which is composed of multiple G-Residuals. The whole backbone network is a full-volume network with a network depth of 24 layers. The two Tiny Hourglass modules adopt the way of intermediate supervision [30]. Take the output characteristic map and input characteristic map of the first Tiny Hourglass module as the input of the second Tiny Hourglass module. The network structure of CenterNet based on Tiny Hourglass24 backbone network is shown in Fig. 4, in which *A*, *B* and *C* are jump connection layers.

3. Training process

3.1. Loss function improvement

*GIOU_* Loss as the loss function of Bounding box, and uses binary cross entropy and Logits loss function to calculate the loss of class probability and target score. The calculation formula is shown as (1) (2) (3):

$$Loss_{coord} = \sum_{i=0}^{S} \sum_{j=0}^{B} l_{ij}^{obj} \left(1 - GIOU_{ij}\right) \tag{1}$$

$$GIOU_{ij} = \frac{J}{U} - \frac{A - U}{A} \tag{2}$$

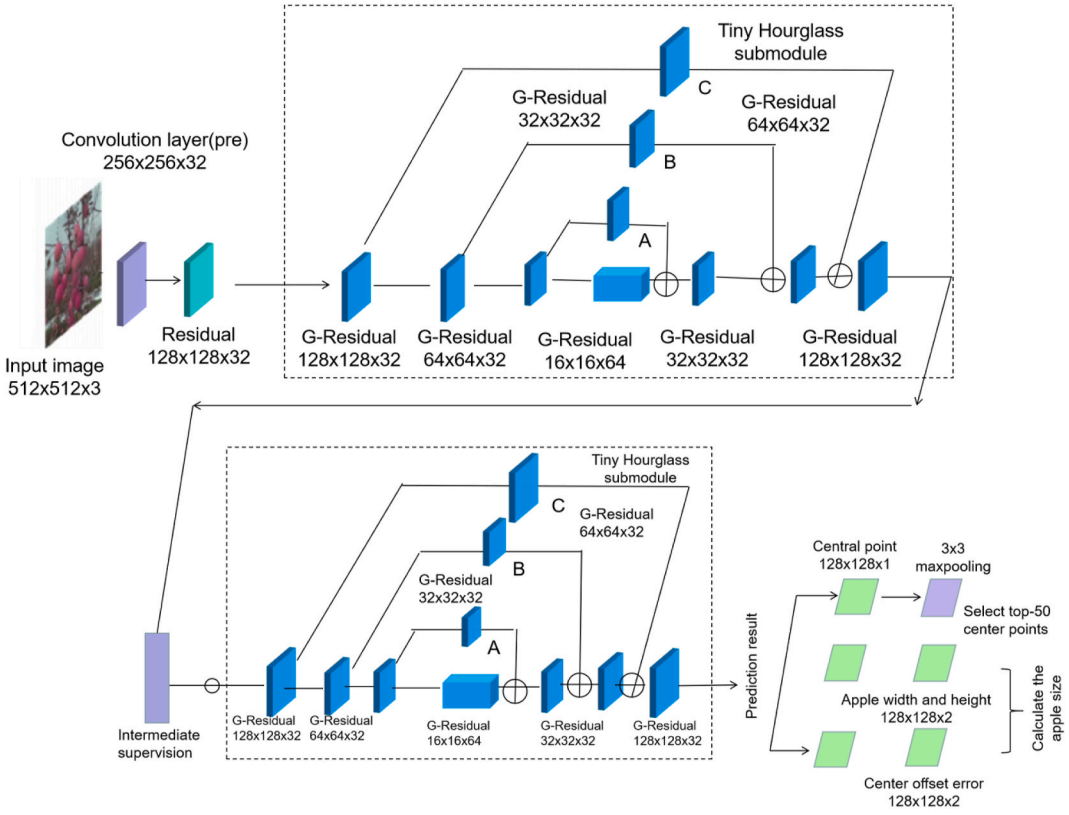


Fig. 4. CenterNet backbone.

$$U = \hat{w}_i \hat{h}_i + w_i h_i - J \tag{3}$$

$Loss_{coord}$-Target position loss function;

l_{ij}^{obj}-A priori box j generated by the cell I contains the target;

J—Intersection area of border; U—Union area of border;

A—Minimum circumscribed rectangular area of the border;

$w_i h_i$—Predict the height and width of the box;

S—Minimum circumscribed rectangular area of real frame and prediction frame;

β—Area where the real box and the prediction box are merged;

$\hat{w}_i \hat{h}_i$— $GIOU\_$ Loss of real border height and width in function;

In this paper, the loss function considering the Euclidean distance of the center point of the prediction frame and the overlap ratio parameter is used as the deviation index of the prediction frame deviation [31]. The deviation indicator is shown as formula (4). The deviation regression process is shown as Fig. 5. The (w^{gt}, h^{gt}) represent the height and width of the prediction box and the real box.

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{4}$$

α—Weight function.

v—Difference square of diagonal inclination angle of rectangle between real box and prediction box.

The objective function is improved as (5):

$$Loss_{coord} = \sum_{i=0}^{S^2}\sum_{j=0}^{B} l_{ij}^{obj}\left(1 - CIOU_{ij}\right) \tag{5}$$

This objective function increases the distance measurement of the center point, which can be direct.

Minimize the distance between two target frames, and the convergence speed is faster than Giou loss. The function is lost, and considering the different situations, the real frame and prediction are avoided. The non-convergence of the frame relation can effectively improve the convergence rate of the object. The recognition rate in the case of occlusion optimizes the relationship between borders.

3.2. Fast ROI identification algorithm for bunched and pluckable apple stalks

Through the global detection of input images and the fusion of multi-scale feature recognition targets, the rapid detection of the Region of Interest (ROI) of Apple bunches and fruit stalks is realized, and the ROI of pluckable fruit stalks are screened out through the connectivity between Apple bunches and corresponding fruit stalks. At present, the identification and location of picking points on fruit stalks are mainly based on the prediction and location of fruit shape characteristics or the identification of fruit stalks according to the relationship between fruit stalks and fruit positions, and then Identify the picking point on the fruit stalk.

The YOLOv5 network structure for target detection of Apple bunches and fruit stalks is shown in Fig. 6. The YOLOv5 model first modifies the input network image into 736 × 416 pixels, and after CSPdarknet-53, it outputs the feature images with three sizes of 92 × 52, 46 × 26 and 23 × 13 pixels respectively. Based on the K-means clustering algorithm, three anchor points with different sizes under each scale feature map are obtained, and three size bounding boxes are predicted for each scale feature map.

3.3. Improved identification method based on YOLOv5

Compared with RGB (Red, Green, Blue), HSV (Hue, Saturation, Value) can express the degree of brightness, vividness and hue of

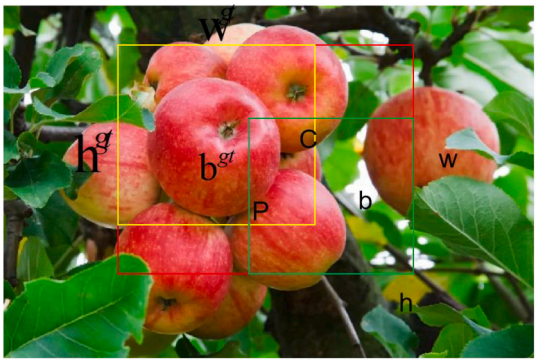


Fig. 5. Loss function GIOU border chart.

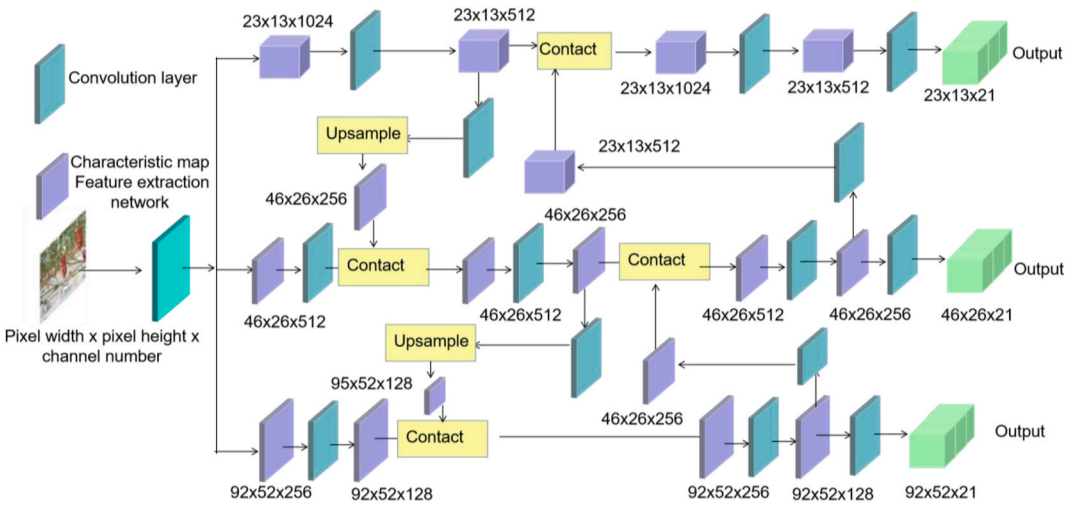


Fig. 6. YOLOv5 model architecture for detecting Apple clusters and stems.

color more intuitively, and its brightness has little influence on color, so it is often used to segment the target of a specified color. The detection frame part of YOLOv5 is extracted from the image, and the RGB image in the frame is converted into HSV image. The conversion formula is shown as formula (6) (7) (8):

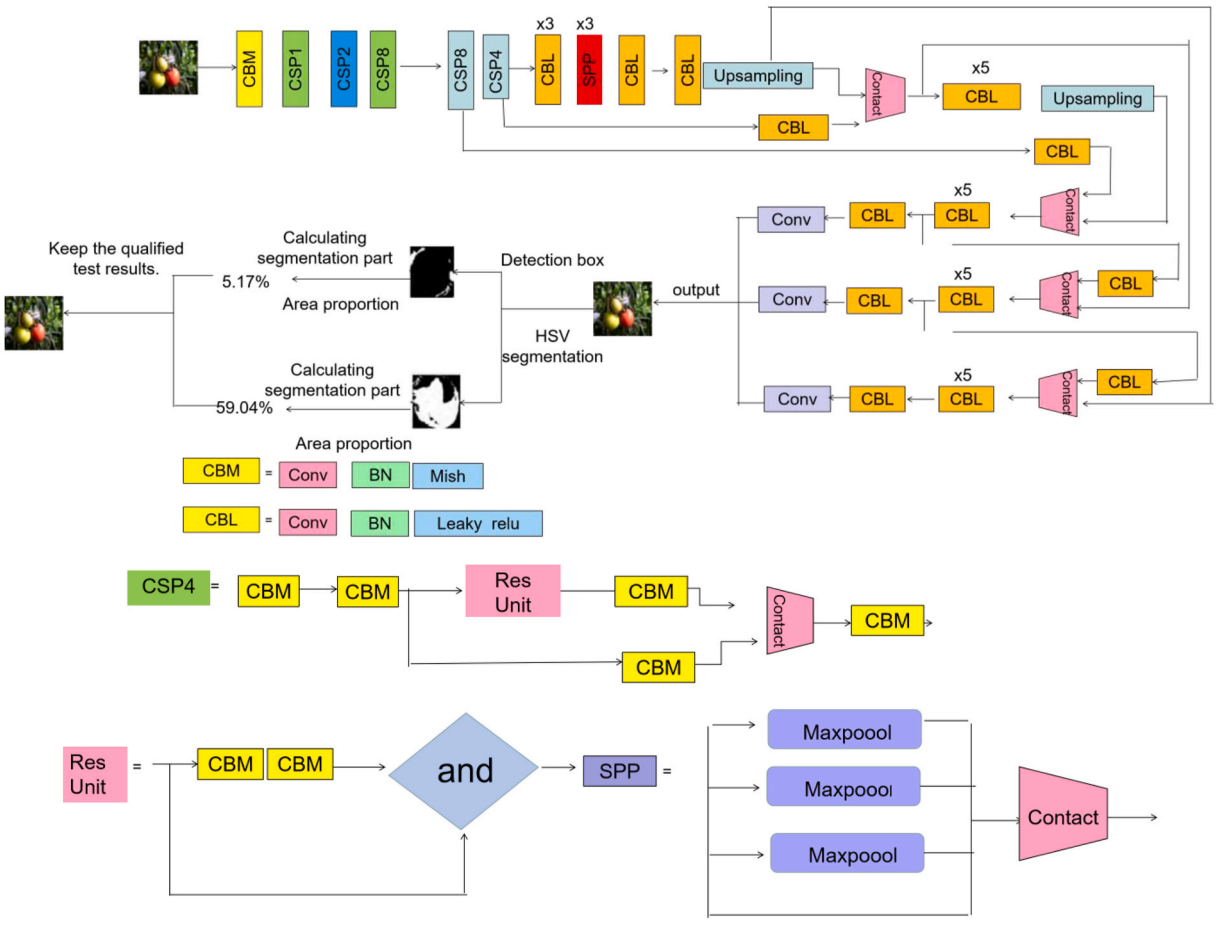


Fig. 7. Recognition process of ripe Apple based on improved YOLOv5 as the backbone in CenterNet.

$$V = \frac{1}{3(R+G+B)} \tag{6}$$

$$S = 1 - \frac{3}{R+G+B}[\min(R,G,B)] \tag{7}$$

$$H = across\left\{\frac{[(R-G)+(R-B)/2]}{\left[(R-G)^2+(R-B)(G-B)\right]^{\frac{1}{2}}}\right\} \tag{8}$$

Where H is the color tone; S is saturation; V is brightness; R is the red value; G is the green value; B is the blue value. The red part in the detection frame is segmented along the H component, and binarized to calculate the area ratio of the segmented part in the detection frame. Because most of the segmented parts are irregular shapes, in order to accurately represent the proportion of segmented parts in the detection frame, this paper uses the number of pixels to represent the area, and determines whether the fruits in the detection frame are ripe Apple by calculating the proportion of segmented parts in the total number of pixels in the detection frame. The pixel ratio calculation is shown as formula (9):

$$A = \frac{\sum i}{w \cdot h} * 100\% \tag{9}$$

Type A is the proportion of red Apple in the detection frame; I is the pixel of the divided part; W and h are the width and height of the YOLOv5 detection frame, respectively. Choosing the right proportion can improve the recognition accuracy, but the proportion setting Excessive size means that the red part of Apple needs to occupy a large part of the detection frame. By comparing the recognition accuracy of different proportions. Missed rate and error rate, select the appropriate proportion as the screening condition, will exceed the excessive proportion is taken as the target and the test result is output, but the proportion is not reached. Target elimination, the specific identification process is shown in Fig. 7. As shown in Fig. 7, the model is mainly composed of modules. SPP (Spatial pyramid pooling) module includes the largest pool layer with different scales, which is used to increase the receptive field.

4. Experiment and analysis

4.1. Comparison of recognition results of Tiny Hourglass network with different depths

The depth of the network has a great influence on the recognition performance. The shallow network will lead to its weak feature extraction ability. Deepening the number of network layers will extract more complex deep features, but it will lead to the reduction of recognition speed. In order to improve the recognition speed of the network without reducing the recognition accuracy, before determining the depth of the backbone network, this experiment designed three kinds of Tiny Hourglass networks with different depths for performance comparison [32], with the set depths of 32, 24 and 12 layers respectively. The recognition results of the networks in test sets A and B are shown in Table 1.

As can be seen from Table 1, with the decrease of the network depth, the F1 value of the network identification decreases, and the average accuracy and F1 value of the network on test set B decrease seriously. When the depth of the backbone network is 24 layers, the average recognition accuracy and F1 value on test set A are 98.90% and 96.39% respectively; Average accuracy and F1 value identified on test set B.

The average recognition time of a single image is 0.069 s, which shows good recognition accuracy and speed on both kinds of test sets. It can be seen that the 12-floor shallow network has a poor effect on fruit recognition in dense scenes. On the test set B, the number of Apples in a single image increases, the size of the fruit in the image decreases, the Apples overlap each other, and the Apples are severely blocked by branches.

In order to verify the recognition performance of the Tiny Hourglass 24 network [33] on Apple Target. The recognition results under two test sets are shown in Table 2.

It can be seen from Table 2 that Tiny Hourglass 24 has the best performance in recognition accuracy and recognition speed. However, because of its shallow network and limited feature extraction ability, the recognition performance on test set B in dense scenes is obviously reduced, and its F1 value and average accuracy are reduced by 4.

Fig. 8 shows the recognition result of CenterNet based on Tiny Hourglass 24, in which the purple box is the Apple recognized by the

Table 1
Backbone network identification results of Different depths.

Test set	Backbone network	average Accuracy/%	F1 value/%	Average of single image Identification time/s
A	Tiny Hourglass-32	98.48	96.86	0.078
	Tiny Hourglass-24	98.90	96.39	0.068
	Tiny Hourglass-12	93.49	94.80	0.065
B	Tiny Hourglass-32	94.10	94.00	0.101
	Tiny Hourglass-24	93.63	92.91	0.069
	Tiny Hourglass-12	74.18	83.78	0.072

Table 2
Detection results of different backbone networks.

Test set	Backbone network	average Accuracy/%	F1 value/%	Average of single image Identification time/s
A	Tiny Hourglass-24	98.90	96.39	0.068
	DLA-34	96.44	95.90	0.103
	ResNet-18	94.95	94.63	0.065
B	Tiny Hourglass-24	93.63	92.91	0.069
	DLA-34	91.16	93.41	0.103
	ResNet-18	81.30	88.28	0.064

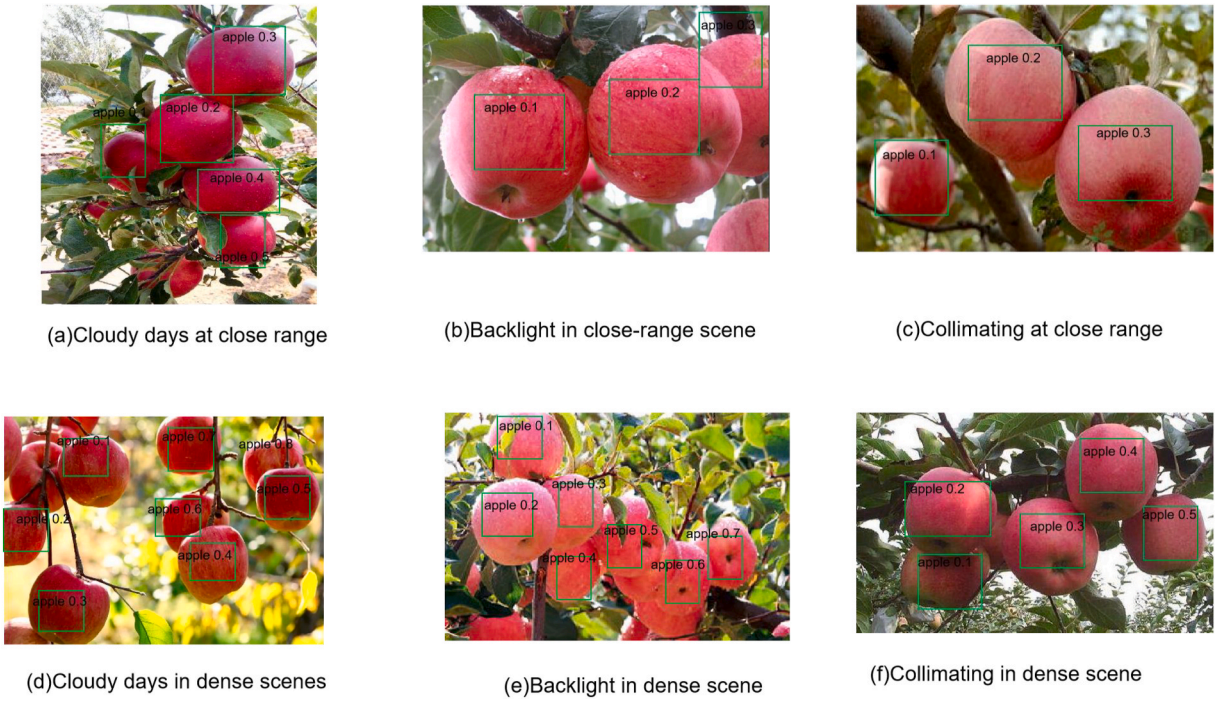


Fig. 8. Detection results of CenterNet
(a) Cloudy days Near object detection results (b) Backlight scene detection results (c) Collimating scene detection results (d) Cloudy days sparse object detection results (e) Backlight scene sparse object detection results (f) Collimating scene Sparse object detection results.

network, and the red box is the Apple missed. Fig. 8 (a)–(c) show the detection results in the near target scenario, which can accurately detect targets except for those with a large range of leaf occlusion. Fig. 8 (e)–(f) show the detection results in sparse target scenes, which can accurately detect targets with incomplete display. As can be seen from Fig. 8, the network has a good recognition effect on the test set An under the conditions of backlighting and sunlighting on cloudy and sunny days, and it can also accurately recognize Apples with serious shading areas. In test set B, the network has a better recognition effect in cloudy environment, and it has a better recognition effect on overlapping and occluded Apples. In the case of backlighting, the surface color of Apples is dark, and there is a small amount of missing recognition in the network when Apples are heavily blocked by branches. Under the condition of smooth light, the surface brightness of Apples is enhanced, and some surface color features become white, which leads to some missed detection of the blocked Apples.

4.2. Comparative experimental analysis of algorithm performance

HSV is used to segment the red area of Apples in the detection frame, and the area proportion of the segmented part in the detection frame is calculated. The lower the proportion, the less the number of missed Apples, but it is difficult to rule out the non-mature period. The total number of target Apples and the number of misidentified Apples are taken as the total, the proportion of identified mature Apples in the total is taken as the accuracy rate, the proportion of misidentified immature Apples is the error rate, and the proportion of undetected mature Apples is the Missing recognition rate. The calculation formula is as follows (10) (11) (12):

$$A = \frac{N_1}{S} \text{x} 100\% \tag{10}$$

$$E = \frac{N_2}{S} \text{x} 100\% \tag{11}$$

$$M = \frac{N_3}{S} \text{x} 100\% \tag{12}$$

where A is the correct rate; E is the error rate; M is the missed detection rate; S is the total number of Apples; N_1 is the number of ripe Apples identified; N_2 is the number of immature Apples misidentified; N_3 is the number of unrecognized ripe Apples. The preliminary analysis shows that when the proportion is greater than 20% and less than 10%, the recognition accuracy rate drops greatly, which seriously affects the recognition results. To further determine the proportion selection, the recognition effect of 10%–20% interval proportion was tested, as shown in Fig. 9. As can be seen from Fig. 9, when the ratio is 16%, and the correct rate is the highest, reaching 94.77%. Considering all the indexes, this paper selects 16% with the highest correct rate as the proportion of Apple recognition algorithm in mature period.

To verify the performance of our method, it is compared with the improved Hough circle transformation algorithm proposed by YOLOv5 by Guo etc. [30] and the improved YOLO v5 algorithm proposed by Zhang [31]. It can be seen from Table 3 that the correct rate of our algorithm is 94.77%, which is 4.30% higher than that before the improvement. The recognition error rate of the improved algorithm is 0.65%, which is 5.29% lower than before. Compared with the original algorithm, the missed detection rate of the improved algorithm increased by 0.99 percentage points, of which 83.33% was the result of fruit color change, and 16.67% was caused by more shading.

To test the practicality of this algorithm, calculate the time it takes to call a camera on different devices to capture images and recognize the first target apple. The camera used in the experiment is Realsense depth camera, and the workstation and micro-industrial computer are each set. Before each set is tested, the position of the target Apple is changed 10 times, and the testing time is shown in Table 4.

4.3. Yov5 android deployment

To verify the reliability of the model, a field test was conducted in the Apple sunlight greenhouse of the North Campus of Northwest. The tests are as follows: ① Deploy the data files of the two network models on Android, and generate the mobile phone detection APP. ② Randomly select areas in the solar greenhouse to collect data. In order to meet the randomness, the image data includes different shapes such as distance, occlusion, etc. ③Manually identify the collected data, and classify the Apple fruits with different colors, whether they are covered or not. ④ The collected image data are identified by using the mobile phone application of two network models, and compared with the manual identification results to analyze the model accuracy. In order to verify the actual detection effect, the mobile phone application generated by improved YOLOv5 and YOLOv5 models are used for field detection, and the detection results are counted. Taking the artificially identified Apple fruit results as reference, the detection results of the two models were compared, analyzed and evaluated. The two models respectively identified the number of red and green Apples, and the ratio of the total identified number to the artificial identified number as the detection accuracy of the two models. The statistical results are shown in Table 5.

From Tables 5 and it can be seen that the recognition accuracy. The total recognition accuracy is increased by 1.6% compared with the YOLOv5 model. The recognition rate of the improved the YOLOv5 model is better than YOLOv5 model in the case of shading or overlapping multiple fruits. In order to verify the effect of model detection in complex environment, the single fruit, multiple fruit, sheltered fruit and uncovered fruit of Apple green fruit and red fruit were statistically distinguished. The results are shown in Table 6.

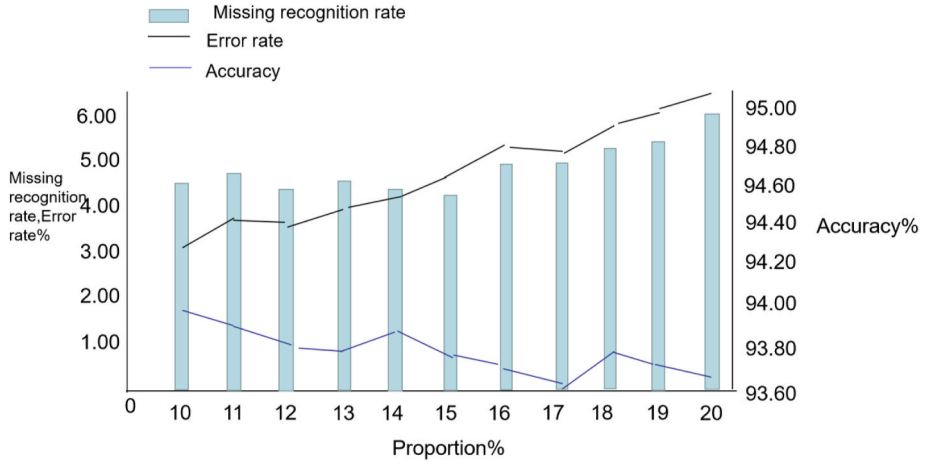


Fig. 9. The recognition effect when the proportion of segmented pixels is 10%–20%.

Table 3
Performance comparison of different algorithms.

Algorithms	Accuracy/%	Error rate/%	Missing recognition rate/%	Detection speed/ms
YOLOv5	90.47	5.94	3.59	22.18
Hough circular transformation	86.82	9.65	3.53	398.00
YOLO v4	91.28	3.49	5.23	54
Ours	94.77	0.65	4.58	25.86

Table 4
Actual detection time of workstation.

Equipment	Group number									
	1	2	3	4	5	6	7	8	9	10
Work station	0.52	0.51	0.51	0.50	0.51	0.49	0.51	0.50	0.52	0.51
Industrial personal computer	1.51	1.50	1.48	1.46	1.48	1.48	1.48	1.48	1.47	1.49

Table 5
Results of all test indexes of two models.

Model	Red fruit precision	Green fruit precision	overall accuracy
Ours	97.6	96.2	96.8
YOLOv5	96.2	94.4	95.2

Table 6
Detection results of Apple fruit in complex environment.

Parameter	Apple green fruit				Apple red fruit			
	simple fruit		Duoguo		simple fruit		Duoguo	
	shelter	Unobstructed	shelter	Unobstructed	shelter	Unobstructed	shelter	Unobstructed
Number of samples/piece	50	50	80	80	50	50	80	80
Recognition rate/%	100	100	96	98	100	100	98	98

The recognition rate of a single fruit can reach 100% with green fruit. Because the color of Apple green fruit is easily confused with leaves, stems, etc. at night, and it is difficult to distinguish boundary conditions when multiple fruits overlap, the recognition rate of Apple green fruit and multiple fruits is slightly lower than that of Apple red fruit. Table 7 shows the results of the ablation comparison experiment.

The loss curve obtained from the improved CenterNet model training is shown in Fig. 10, which is compared with the original YOLOv5 model.

5. Conclusion

This proposed method can effectively detect apples even at night and under insufficient lighting. The center network based on the minute hourglass 24 has more advantages than the recognition method based on the anchor frame and the recognition method based on multiple key points by predicting the center point of the target. The whole recognition process does not use anchor frames and NMS post-processing, which reduces network parameters and is more suitable for multiscale apple target recognition in dense scenes. In the future, the structure can be further adjusted to expand the learned dataset to achieve better detection results. Moreover, it is possible to connect the camera to an embedded platform and collect and process image or video data in real-time in real traffic environments to verify the generalization ability of the model. It is also possible to perform lightweight processing on proposed networks with higher width and depth.

Ethics declarations

Review and/or approval by an ethics committee was not needed for this study because this article does not contain any studies with human participants or animals performed by any of the authors.

Data availability statement

Data presented in this work can be made available on request from the corresponding author.

Table 7
Results of ablation comparison experiments.

Network model	Enter the image size.	model parameter	amount of calculation(G)	mAP @0.5:0.95	mAP @0.5
YOLOv5s	608 × 608	7022326	14.3	0.689	0.893
MBConvBlovk + YOLOv5	608 × 608	3341858	5.8	0.634	0.882
Stem + MBConvBlovk + YOLOv5	608 × 608	3343730	6.5	0.656	0.906
CBAM + MBConvBlovk + YOLOv5	608 × 608	3354456	5.9	0.648	0.884
CAM + Stem + MBConvBlovk + YOLOv5	608 × 608	3356034	6.5	0.662	0.914
SAM + Stem + MBConvBlovk + YOLOv5	608 × 608	3344024	6.5	0.659	0.908
CBAM + Stem + MBConvBlovk + YOLOv5	608 × 608	3356328	6.5	0.669	0.915
ECA + Stem + MBConvBlovk + YOLOv5	608 × 608	3343745	6.5	0.659	0.904
ECBAM + Stem + MBConvBlovk + YOLOv5	608 × 608	3344039	6.5	0.666	0.916

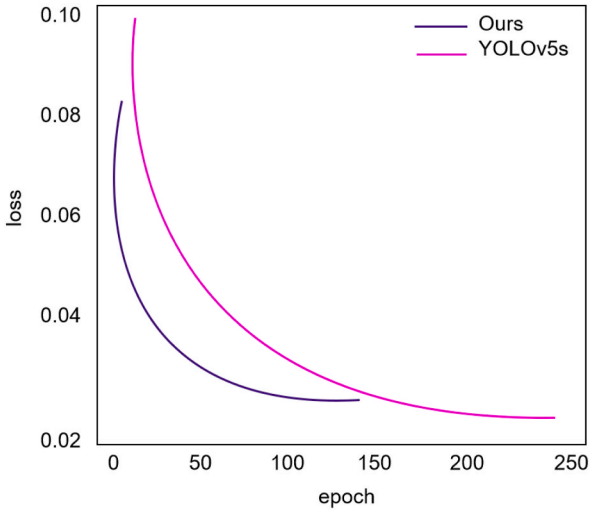


Fig. 10. Loss curve.

CRediT authorship contribution statement

Han Zhou: Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Haotian Liu reports financial support, administrative support, article publishing charges, equipment, drugs, or supplies, statistical analysis, travel, and writing assistance were provided by College of Science and Engineering, The University of Edinburgh, Edinburgh. Haotian Liu reports a relationship with College of Science and Engineering, The University of Edinburgh, Edinburgh that includes: board membership, consulting or advisory, employment, equity or stocks, funding grants, non-financial support, paid expert testimony, speaking and lecture fees, and travel reimbursement.

Acknowledgements

This work was supported by the Education Department of Hainan Province, project number: Hnky2022ZD-20.

References

[1] Yu Qiu, Junying Han, Feng Chengzhi, Yongwei Chen, Identification of Apple cultivars based on convolutional neural network [J], Comput. Mod. (12) (2021) 65–71.
[2] Mengran Zhou, Jingjing Yan, Wenhao Lai, Jinguo Wang, Hu Feng, Kai Bian, Qianqian Kong, Multi-spectral identification of damaged Apples based on complete local binary pattern [J], Journal of Food Safety and Quality Inspection 12 (23) (2021) 9086–9092, https://doi.org/10.19812/J.
[3] Ling Ma, Research on the Identification Method of Apple Leaf Disease Based on Deep Learning [D], Xijing University, 2021, https://doi.org/10.27831/D.cnki.gxjxy.20008.00000000074.
[4] Yongsheng Si, Shanshan Cao, Xiaoxue Zhang, Ji Ying, Jixing Lu, Identification of Apple bitter pox and bump injury based on CT images [J], Journal of Agricultural Machinery 52 (10) (2021) 377–384.

[5] Lijuan Liu, Dou Peipei, Shine Wong, Study on the method of image recognition of overlapping and occluded Apples in natural environment [J], Chinese Journal of Agricultural Machinery Chemistry 42 (6) (2021) 174–181, https://doi.org/10.13733/j.jcam.issn.2095-5553.2021.
[6] X. Yu, X. Ye, S. Zhang, Floating pollutant image target extraction algorithm based on immune extremum region, Digit. Signal Process. 123 (2022) 103442.
[7] Xiancheng Ren, Research on Identification Method of Apple Based on Hyperspectral Technology and Chemometrics [D], Tarim University, 2021, https://doi.org/10.27708/d.cnki.gt.lmd.20008.000000000005.
[8] X. Xie, G. Cheng, J. Wang, et al., Oriented R-CNN for object detection[C], Proceedings of the IEEE/CVF International Conference on Computer Vision (2021) 3520–3529.
[9] S.K. Pal, A. Pramanik, J. Maiti, et al., Deep learning in multi-object detection and tracking: state of the art[J], Appl. Intell. 51 (9) (2021) 6400–6429.
[10] D. Wang, D. He, Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning[J], Biosyst. Eng. 210 (2021) 271–281.
[11] X. Zhou, G. Sun, N. Xu, et al., A method of modern standardized apple orchard flowering monitoring based on S-YOLO[J], Agriculture 13 (2) (2023) 380.
[12] H. Mirhaji, M. Soleymani, A. Asakereh, et al., Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions[J], Comput. Electron. Agric. 191 (2021) 106533.
[13] Wei Yahui, Gengnan Huang, Fupei Wu, Apple recognition method based on improved watershed algorithm [J], Packag. Eng. 42 (8) (2021) 255–260, https://doi.org/10.19554/j.cnki.1001-3563.2021.08.
[14] Yuliang Jiang, Research on the Method of Apple Target Recognition and Location under the Complicated Background Based on OpenCV [D], harbin university of science and technology, 2021, https://doi.org/10.27063/D.
[15] Li Dahua, Bao Xuejuan, Yu Xiao, Gao Qiang, Detection and identification of green Apples in natural environment based on YOLOv3 network [J], Laser J. 42 (1) (2021) 71–77, https://doi.org/10.14016/j.cnki.jgzz.2021.01.
[16] Jian Wang, Xuehua Liu, Pathological identification of Apple leaves based on depth separable convolution, Computer System Application 29 (11) (2020) 190–195.
[17] Huan Lei, Zeyu Jiao, Jingqi Ma, Liangsheng Wu, Zhenyu Zhong, Fast Apple variety identification algorithm based on multi-feature fusion and SVM [J], Automation and Information Engineering 41 (4) (2020) 13–17.
[18] X. Yu, X. Ye, S. Zhang, Floating pollutant image target extraction algorithm based on immune extremum region, Digit. Signal Process. 123 (2022) 103442.
[19] Jinglu Jie, Yuan Ma, Jiaxin Wu, Ke Chen, Design of image recognition system for rotten Apples based on deep learning [J], Journal of Nanyang Institute of Technology 12 (4) (2020) 66–70, https://doi.org/10.16827/j.cnki.41-1404/z.27.
[20] Y. Tian, G. Yang, Z. Wang, et al., Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J], Comput. Electron. Agric. 157 (2019) 417–426.
[21] Youjun Yue, Tian Bokai, Hongjun Wang, Hui Zhao, Application of improved VGG model in Apple appearance classification [J], Sci. Technol. Eng. 20 (19) (2020) 7787–7792.
[22] A. Kuznetsova, T. Maleva, V. Soloviev, YOLOv5 versus YOLOv3 for Apple detection[M]//Cyber-Physical Systems: Modelling and Intelligent Control, Springer International Publishing, Cham, 2021, pp. 349–358.
[23] Jing Xiaomei, Research on the Algorithm of Apple Target Recognition in Natural Environment [D], Xi 'an University of Science and Technology, 2020.
[24] P. Mondino, J.L. Gonzalez-Andujar, Evaluation of a decision support system for crop protection in apple orchards[J], Comput. Ind. 107 (2019) 99–103.
[25] Jiayu Dai, Method of Identifying Green Apples at Night by Apple Picking Robot [D], Zhejiang University of Technology, 2020.
[26] S. Qummar, F.G. Khan, S. Shah, et al., A deep learning ensemble approach for diabetic retinopathy detection, IEEE Access 7 (2019) 150530–150539.
[27] Jing Weibin, Hu Haitang, Cheng Cheng, Li Cunjun, Jing Xia, Zhijun Guo, Recognition and counting of ground Apples based on deep learning [J], Jiangsu Agric. Sci. 48 (5) (2020) 210–219, https://doi.org/10.15889/J.ISSN.1002-05.
[28] X. Yu, X. Tian, A fault detection algorithm for pipeline insulation layer based on immune neural network, Int. J. Pres. Ves. Pip. 196 (2022) 104611.
[29] Z. Rehman, M.A. Khan, F. Ahmed, et al., Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: an application for smart agriculture[J], IET Image Process. 15 (10) (2021) 2157–2168.
[30] G. Guo, Z. Zhang, Road damage detection algorithm for improved YOLOv5[J], Sci. Rep. 12 (1) (2022) 1–12.
[31] Shifu Zhang, Research on Apple Target Recognition and Location Algorithm Based on Deep Learning [D], Zhejiang University of Technology, 2020.
[32] Jing Weibin, Li Cunjun, Jing Xia, Zhao Ye, Cheng Cheng, Apple tree side view fruit identification based on deep learning, China Agricultural Information 31 (5) (2019) 75–83.
[33] Yuangang Zheng, Research on deep neural network algorithm for apple image recognition [J], Informatization Construction (8) (2019) 59–60.