Improved Deep Convolutional Neural Network-Based Method for Detecting Winter Jujube Fruit in Orchards

Tianzhen Liu, Yingchun Yuan*, Guifa Teng, Xi Meng

Abstract—**To address the issue of rapid and precise classification and recognition of winter jujube fruit in orchards, we propose an enhanced deep convolutional neural network and an improved loss function recognition method based on the YOLOv4 model. The multi-scale feature fusion Cross Stage Partial Connections (CSP) structure and Convolution Block Attention Module (CBAM) are used to enhance the model, minimise the network scale, amplify the feature extraction capability and improve the false detection in fruit classification as well as the missing detection of occlusions. The Softmax cross-entropy loss function replaced the Sigmoid binary cross-entropy loss function as the classification loss function. Additionally, the EIoU loss function was introduced to replace the CIoU loss function as the bounding box regression loss function. These changes were made to further reduce false detection in fruit classification and enhance the accuracy of prediction frames. The experimental results indicated that *P (Precision), mAP (Mean Average Precision)* and *IoU* (*Intersection over Union*) values for three classes of winter jujube fruits were 81.86%, 82.46% and 81.35%, respectively. The model has 26.9M parameters, a size of 108MB and a detection speed of 28.8 frames per second. This method boasts high accuracy and robustness, providing significant reference value for accurately detecting winter jujube fruit in orchards.**

Index Terms—**convolutional neural network, fruit detection, loss function, winter jujube**

I. INTRODUCTION

WINTER Jujube plants are extensively cultivated across a wide area in China, primarily in regions such as Hebei, Shandong and Shaanxi. High-quality winter jujube fruits are highly favoured by consumers, bringing significant economic benefits to the local area[1]-[2]. The flavour and quality of winter jujube fruit are closely tied to its maturity, which is determined by the colour of the pericarp. The fruit, from the 1/3 red stage to the semi-red stage, possesses the best overall quality, with good appearance, taste, sugar and nutritional content. It can be identified as the optimal harvesting period [3]. Currently, the harvesting of winter jujube is primarily done manually. Harvesting winter jujube fruits during the optimal period is labour-intensive and inefficient. Hence, studying a detection method for winter jujube fruits of varying maturity levels in orchards is of immense practical importance for realising mechanized picking of these fruits.

The identification of winter jujube falls under the category of fruit recognition. Numerous solutions have been proposed based on conventional machine learning methods. **Wang** et al. [4] proposed an image segmentation method based on RGB colour space for recognizing the maturity of Lingwu Long Jujube. The accuracy rate of maturity level recognition following segmentation exceeds 90%. **Bi** et al. [5] proposed a citrus visual recognition model using a multiple segmentation method, achieving an average accuracy of 86.6%. **Liu** et al. [6] used the chosen maturity evaluation index and spectral index to construct a multiple linear regression evaluation model, thereby achieving accurate classification. **Zhao** et al. [7] developed a spectral classification model using spectral data to ascertain the maturity of apples for picking. Among them, the SIQI + SVR classification model demonstrated optimal discrimination accuracy, with the prediction set achieving a high accuracy rate of 85.71%.

None of the aforementioned research results addressed the detection speed and efficiency. Some of the results used spectral detection, which is not only challenging to deploy and implement in the picking device but also susceptible to the influence of natural light, leading to significant errors and reduced robustness. It is evident that numerous issues exist in the fruit recognition classification method based on traditional machine learning, making it challenging to meet the demands of real-world scenarios.

In recent years, deep learning methods have increasingly been used in the field of agricultural fruit recognition. Research on winter jujube fruit identification primarily includes the following problems. **Lu** et al. [8] used an algorithm that combines YOLOv3 and manual features to design a winter jujube classification robot. This robot classifies the fruit based on the red proportion of its surface into three categories: suitable for long-term storage, short-term storage and sales. This allows for the efficient identification and classification of harvested winter jujubes. However, this method examined the fruits post-harvest in a laboratory setting and did not conduct recognition tests on winter jujube fruits in their natural environment. **Wang** et al.

Manuscript received July 19, 2023; revised January 2, 2024.

This research work was funded by Science Research Project of Hebei Education Department (ZC2024096), the Natural Science Foundation of Hebei (F20204003), and the Science and Technology Plan Project of Baoding, P.R.China (2211ZG010).

Tianzhen Liu is an Associate Professor at the College of Information Engineering, Baoding University, Baoding, Hebei, China, and a PhD student of College of Information Science and Technology, Hebei Agricultural University, Baoding, Hebei, China. (e-mail: liutz75@163.com).

Yingchun Yuan* is a Professor of the College of Information Science and Technology, Hebei Agricultural University, Baoding, Hebei, China. (corresponding author to provide phone: +086-0312-7521530; fax: +86 -0312-7521530; e-mail: nd_hd_yyc@163.com).

Guifa Teng is a Professor of Hebei Key Laboratory of Agricultural Big Data, Baoding,Hebei, China. (e-mail: 1847412438@qq.com).

Xi Meng is a professor at the College of Future Information Technology, Shijiazhuang University, Shijiazhuang, Hebei, China. (e-mail: 1969983871 @qq.com).

[9] identified winter jujube at full green maturity and red maturity with insufficient sample quantity and unbalanced categories in a natural environment, using data augmentation and Faster RCNN. The model's generalization effect was validated by comparing it with YOLOv3, but the fruit classification did not meet the actual harvesting requirements of winter jujube. **Liu** et al. [10] incorporated the SE block into the YOLOv3 model to detect and recognize winter jujubes in natural scenes. The recognition performance improved under varying conditions of density, light and maturity. However, the model did not classify and identify the winter jujube fruit. The aforementioned studies on winter jujube fruit identification struggle to fulfil the requirements for detecting these fruits in a natural environment.

Numerous studies have been conducted on the identification of various other types of fruits. Among them, **Zhao** et al. [11] enhanced the YOLOv3 network by altering the backbone network and implementing a combined loss function of Focal Loss and CIoU Loss. This improved the model's identification accuracy, enabling the recognition of mature and unripe apples in an orchard setting. **Zhao** et al. [12] constructed a 13-layer backbone network based on YOLOv3. They used two anchor boxes of different scales to detect and identify three classes of apples, mature, unripe and bagging, in complex environments. They achieved high accuracy and a significant *Intersection over Union (IoU)* mean. **Cao** et al. [13] developed a lightweight YOLOv4-LightC-CBAM model derived from the YOLOv4 model by decreasing the network's width, reducing the number of backbone network layers, simplifying the neck network and incorporating the CBAM block. This model was used for rapid and precise identification of mango fruits in harvesting scenarios. **Wang** et al. [14] used the CSP Darknet53-Tiny network model as the backbone network based on the YOLOv4-Tiny network. They incorporated the CBAM block into the FPN structure of the YOLOv4-Tiny network, enhancing the identification of three classes of blueberry fruits: ripe, semi-ripe and unripe. **Huang** et al. [15] proposed a citrus detection method using the YOLOv5 model, incorporating the CBAM block into the model and replacing the GIoU loss function with the α-*IoU* loss function for bounding box regression. This not only enhanced the average accuracy but also decreased the model size and reduced the detection time. **SUN** et al. [16] enhanced GoogLeNet by implementing category balance loss to boost the detection accuracy of Yulu pear, particularly with the unbalanced quantity of intact, rotten and rust spot conditions. **Zhang** et al. [17] developed a detection model for small target cherries by modifying the backbone based on the YOLOv5 model. They incorporated a Transformer module, converted the PAFPN structure into a BiFPN structure for bidirectional weighted fusion in the neck, and added a shallow downsampling module in the Head. The deployment to the Android platform has significantly enhanced the detection of cherry fruit in field environments.

The aforementioned studies on fruit recognition have achieved enhanced efficiency and results by refining the network model, incorporating the attention mechanism and adjusting the loss function. To address the issues of slow detection speed and subpar detection results in winter jujube fruit detection in orchards, this study employs the YOLOv4 model. It integrates the CSP (Cross Stage Partial Connections) block and the CBAM block, which incorporate multi-scale features and a loss function. Consequently, the CC-YOLO method is proposed for the detection of winter jujube fruits. In conjunction with enhancing the loss function for optimal results, it offers a method for the swift and precise classification and identification of winter jujube fruits.

II. DATASET PRODUCTION

The images in this paper were collected from the Jujube Tree Education Base in Cangxian County, Hebei Province, using a Canon digital camera and a smartphone. The date ranges from late August to early October. The winter jujube fruits are at different stages of growth. Considering the negative impacts of low light and the restricted picking range, the image capture is maintained near the picking target where the light is adequate.

In this experiment, 1000 images were chosen as the dataset. The image annotation tool LabelImg was used to classify and annotate the targets of winter jujube fruits. According to the recommendations of jujube experts and relevant research [4], winter jujube fruits ready for harvest are categorised into three classes, each corresponding to a different maturity level. Stage 1 represents the phase of awaiting harvest, encompassing green ripe, white ripe and red fruit (characterised by a small amount of scattered red on the surface). Stage 2 represents the optimal picking stage, characterised by continuous red and less than 50% red ratio on the surface. Stage 3 is the class that has surpassed the optimal picking period and requires immediate harvesting, encompassing more than 50% of fully ripe red fruit. The three classes are denoted as S1, S2 and S3. Considering that the picture contains both distant and close-up scenes, fruits that are either fully visible or obscured by less than 30% were labelled. Small fruits and those unsuitable for picking due to their distance were discarded. The object boxes were labelled with utmost accuracy, and XML files were subsequently generated.

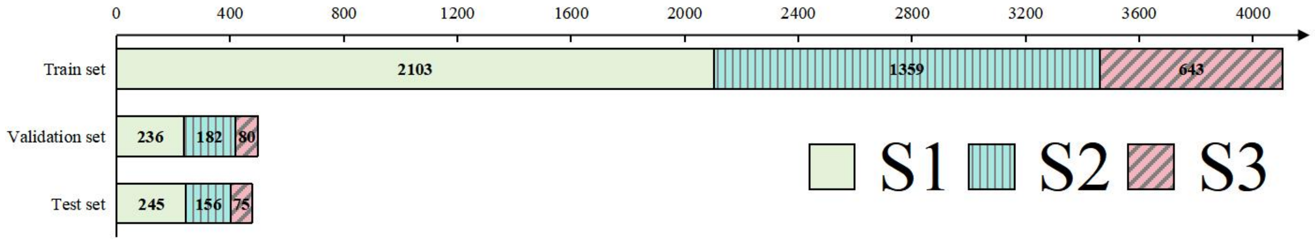


Fig. 1. Quantities of different classes of winter jujube fruit in datasets

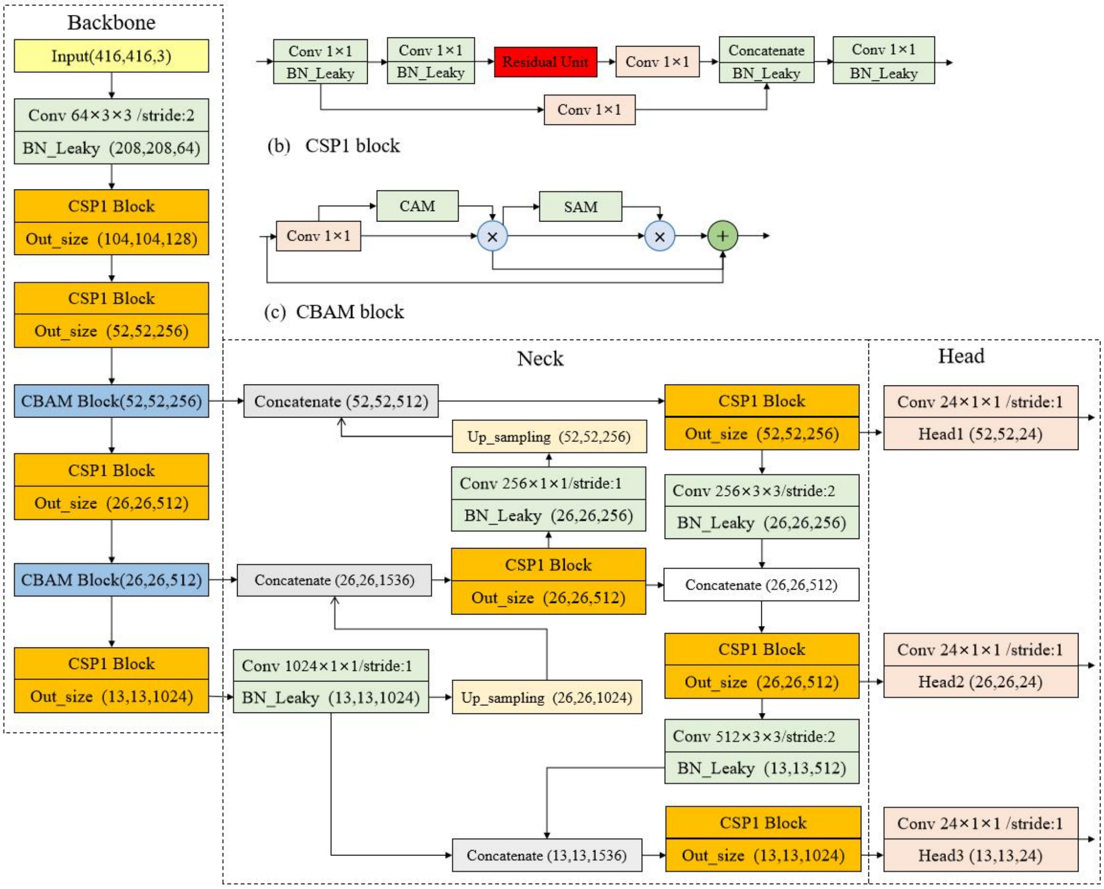
The labelled dataset was randomly partitioned into a training set, validation set and test set at a ratio of 8:1:1. The details of the datasets are shown in Fig. 1. The subtotal of winter jujube fruits in each dataset was 4105, 498 and 476, respectively. Of the three classes, 2584 were in S1, 1697 in S2 and 798 in S3.

III. PROPOSED METHOD

A. Network

The YOLO series models, which are end-to-end deep convolutional neural networks, are extensively used for fruit detection. YOLOv4 [18] is a commonly used model [19]-[20]. It comprises three parts: Backbone, Neck and Head. The backbone primarily consists of five residual units with a CSP structure and one CBM block. The CSP structure is derived from CSP Net [21]. The fundamental concept involves dividing the feature map into two sections, with each section learning the features differently before finally merging them [19]. This mitigates the issue of gradient feature reuse, decreases the computational load and enhances the model's learning capacity. The number of residual units in the five CSP blocks varies, specifically 1,2,8,8,4. This significantly increases the number of model layers and enhances the feature extraction capability. However, the deeper the network is , the better the fit will be, it will fall into local optimum and fail to obtain a better solution. When the number of layers reaches a specific number, increasing the number of layers cannot improve the recognition accuracy of the neural network [13]. Simultaneously, considering the constraints of computational power, reducing the number of layers and constructing lightweight neural networks can often achieve higher detection speed and meet the requirements of recognition accuracy [12]–[15].

The enhanced model is referred to as CC-YOLO, shown in Fig. 2(a). It comprises the CBL, CSP and CBAM blocks. These are located in the backbone. Following the CBL block, the Backbone of the model is constructed using four CSP blocks. The structure of CSP block is shown in Fig. 2(b). The number of residual unit in the CSP block is 1. After the Concat splicing, a batch normalisation layer and a LeakyReLU activation layer are added to expedite network convergence. These are then applied to the neck network, replacing the CBL block in front of the Head layer and further enhancing the feature extraction capability in the neck network.In the enhanced model, the attention mechanism is incorporated to bolster the extraction of effective features and optimise recognition performance. The CBAM [22] block is composed of CAM (*Channel Attention Module*) and SAM (*Spatial Attention Module*), which can effectively enhance the feature extraction capability without significantly increasing the number of computations and parameters. In this paper, CBAM blocks of multi-scale feature fusion method is used for enhancement shown in Fig. 2(c). On the one hand, the feature weights of space and



(b) CSP1 block

(c) CBAM block

(a) CC-YOLO model

Fig. 2. Structure of the improved model

channel are extracted and weighted with the original feature. On the other hand, only the features extracted by the CAM block are weighted with the original feature. Finally, the original feature is directly fused with the first two weighted features. Although a small number of feature dimensions are added, the computational complexity remains low, and the semantic information of features is enhanced, thereby improving the feature extraction ability.

B. Loss function

The precision of fruit positioning significantly influences the success rate of harvesting, which requires a detection model that not only boasts high accuracy in fruit detection but also excels in determining the fruit's width, height and location accurately. To enhance the accuracy of model classification detection, the prediction box size and the precision of the position, it is necessary to further improve the loss function.

The YOLOv4 loss function comprises three components: object confidence loss, classification loss and bounding box loss. The first two losses use the binary cross-entropy loss function, while the latter employs the CIoU loss function.

The CIoU loss function takes into account the overlapping area between the ground truth box and the predicted box, the distance between the centre points of the two boxes and the aspect ratio for location regression, as depicted in formula (1).

$$\mathcal{L}_{CIoU}= 1\text{-}IoU+\frac{\rho^2(b,b^{gt})}{c^2}+\alpha\upsilon \tag{1}$$

$$\alpha=\frac{\upsilon}{(1-IoU)+\upsilon}$$

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

However, when the shape of the winter jujube fruit is nearly round and the aspect ratio is approximately 1, CIoU loss degenerates to DIoU loss. This results in an invalid penalty term, and the prediction box cannot accommodate the ground truth box.

The EIoU loss function is presented as depicted in formula (2). EIoU [23] modifies the aspect ratio penalty term to the difference between the length and width of the ground truth box and the predicted box. This solution addresses the issue of penalty term degradation when the length and width of the winter jujube are similar, thereby enhancing the accuracy of the predicted box.

$$\mathcal{L}_{EIoU}= \mathcal{L}_{IoU}+\mathcal{L}_{dis}+\mathcal{L}_{asp}=1\text{-}IoU+\frac{\rho^2(b,b^{gt})}{c^2}+\frac{\rho^2(w,w^{gt})}{C_w^2}+\frac{\rho^2(h,h^{gt})}{C_h^2} \tag{2}$$

The YOLOv4 model employs the Sigmoid binary cross-entropy loss function as a classification loss, as depicted in formula (3), taking into account both positive and negative sample losses. When the number of positive and negative samples is imbalanced, particularly when the positive samples are few and the negative samples are numerous, the model tends to focus more on the loss of negative samples and overlooks the loss of positive samples. This is not beneficial for enhancing the classification accuracy of the samples. This method is more suitable for binary classification problems where classes are not mutually exclusive.

From Fig. 1 shown above, it is evident that there is a significant imbalance in the quantity of three classes of winter jujube fruit. This is primarily due to the fact that winter jujube is a fresh variety with a non-concentrated maturity period, resulting in a substantial variation in the proportion of each class. During training, the network tends to pay more attention to classes with a large number of samples while diminishing the importance of classes with fewer samples. This causes the recognition performance of the class with a small sample proportion to decline. Furthermore, the three classes of winter jujube fruits are mutually exclusive, making it inappropriate to use the Sigmoid binary cross-entropy loss function.

$$\mathcal{L}_{\mathrm{BCE}}=-\left[\mathcal{Y}*\ln\widehat{\mathcal{Y}}+(1-\mathcal{Y})\ln(1-\widehat{\mathcal{Y}})\right] \tag{3}$$

$$\mathcal{L}_{\mathrm{CE}}=-\mathcal{Y}\ln\widehat{\mathcal{Y}} \tag{4}$$

The Softmax cross-entropy loss function, as depicted in formula (4), concentrates solely on the loss of positive samples, disregarding the loss of negative samples. This approach can directly enhance the classification accuracy of positive samples. It can enhance classification accuracy by addressing the issues of mutual exclusivity among multiple classes and quantity imbalance. Therefore, the classification loss modifies the Sigmoid binary cross-entropy loss function to the Softmax cross-entropy loss function to achieve further improvement.

C. Model parameters and size

The comparison of model network parameters and model size is shown in Fig. 3. CC-YOLO model has only 26.9 M parameters, which is 34.5 M less than the YOLOv4 model and 31.8 M less than the YOLOv3 model. The model is 108 MB, which is only 45.96% and 44.26% of the size of the YOLOv3 and YOLOv4 models, respectively. The data indicate that CC-YOLO is smaller, which will reduce both training and testing time, thereby improving efficiency.

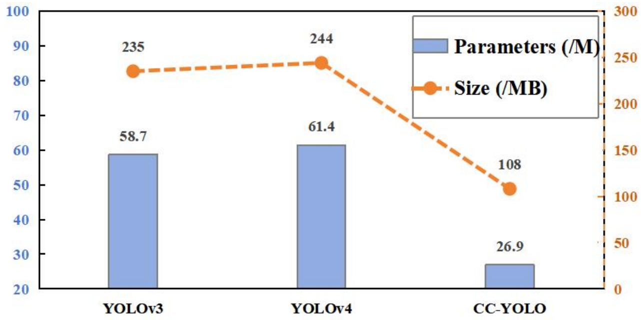


Fig. 3. Parameters and size of the models

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experiment operating environment

All experiments in this paper were conducted on the same server, equipped with an Intel Core i5-10400F CPU at 2.90 GHz, an NVIDIA GeForce RTX 2080SUPER GPU and 16 GB of RAM. The software environment uses a 64-bit Windows 10 system and the TensorFlow deep learning framework.

B. Model training and detection

The input image size is 416 × 416 pixels during model training. The SGD optimiser was used. The batch size is set at 8, and the training set comprises 800 images; thus, each epoch involves 100 iterations. During training, a 50-epoch pre-training was conducted by freezing the backbone network. The initial learning rate was set to 0.001, and the decay rate was set to 0.94. Subsequently, the entire model was trained post-freezing, with the initial learning rate reset to 0.0001. To prevent overfitting, a limit of 10 epochs is set without a decrease in loss value.

The same training and validation sets were used to train YOLOv3, YOLOv4 and CC-YOLO, respectively. The loss value of CC-YOLO decreased rapidly before the 80th epoch, then slowly declined to a minimum value of 1.29 at the 177th epoch. Meanwhile, the YOLOv3 model reached its minimum value of 3.11 at the 196th epoch, and the YOLOv4 model reached its minimum value of 2.92 at the 190th epoch, respectively.

C. Evaluation indicators

In this section, the detection results are evaluated based on the accuracy of the predicted samples and the precision of the prediction box in a pick-oriented situation. The model's primary evaluation metrics are *P* (*Precision*), *mAP* (*Mean Average Precision*) and *IoU* (*Intersection over Union*). *P* represents the ratio of true positive samples to all predicted samples, serving as a measure of the accuracy of the predicted samples. *mAP*, on the other hand, is the mean average prediction across all classes used to gauge the prediction accuracy of all classes. *IoU* is the overlap ratio between the predicted bounding box and the ground truth box, which is used to gauge the accuracy of the predicted location and dimensions. The computation of each index is depicted in formulas (5)–(7) below.

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$mAP = \frac{1}{\mathrm{C}}\sum_{\mathrm{k}=1}^{\mathrm{C}} AP_k \quad (6)$$

$$IoU = \frac{\mathrm{area}(B_p \cap B_{gt})}{\mathrm{area}(B_p \cup B_{gt})} \quad (7)$$

In (5), *TP* represents the number of true positive samples, i.e. the number of fruits correctly detected. *FP* refers to the number of false positive samples, namely, the number of fruits incorrectly detected. The total number of predicted fruits is the sum of the two numbers. In formula (6), C is the number of the classes, and AP_k is the k-class average precision. In formula (7), B_p and B_{gt}, respectively, represent the prediction box and the ground truth box.

Fruit recognition oriented towards picking requires high accuracy in prediction box calculations. Thus, the *IoU* threshold is raised to 0.75 in model detection, and comparative experiments are conducted on model accuracy.

D. Detection speed

The detection speed of the model is determined by the time taken from the input of the image into the network to the output of the detection result. In this paper, we compare the detection and recognition speeds of YOLOv3, YOLOv4 and CC-YOLO using the same test set and images with a resolution of 1024*1024 pixels. The unit of detection speed is measured in FPS (*Frames Per Second*). This means the number of images detected per second. The results are shown in Fig. 4. The 100 images in the test set vary in size, with most being high-resolution. The smallest image measures 800 × 600 pixels, while the largest is 3120 × 2496 pixels. Each model detects and identifies the test set in batches. CC-YOLO detects at a speed of 20.6 FPS, while YOLOv4 operates at 14.3 FPS and YOLOv3 at 15.7 FPS. CC-YOLO runs at a speed 6.3 FPS faster than YOLOv4 and 4.9 FPS faster than YOLOv3.

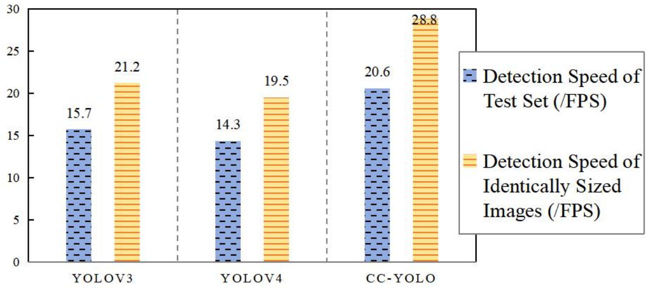


Fig. 4. Comparison of detection speeds among the models

When using high-resolution images of the same size as input for batch detection, the model's detection speed can reach 28.8 FPS. This is 7.6 FPS and 9.3 FPS faster than YOLOv3 and YOLOv4, respectively. The advantage of detection speed is evident, and it can satisfy the real-time requirements at a higher resolution.

E. Overall results of detection

The models identify the fruits in the test set, with the evaluation metrics displayed in Fig. 5-7.

1) Detection precision

P_1, P_2 and P_3 represent the precision values of S1, S2 and S3, respectively. Similarly, AP_1, AP_2 and AP_3 denote the average precision values of S1, S2 and S3, respectively. In Fig. 5, P_1, P_2 and P_3 of CC-YOLO without the improved loss function are 81.23%, 82.55% and 80.24%, respectively. These values are 4.12%–6.34% and 4.33% – 5.25% higher than those of the YOLOv3 and YOLOv4 models. The average of *P* was 81.34%, which was 5.13% and 4.82% higher than YOLOv3 and YOLOv4, respectively. P_2 showed the greatest increase, registering 6.34% higher than YOLOv3 and 5.52% higher than YOLOv4. The model's overall recognition accuracy has also shown significant improvement. The model can accurately detect and distinguish the maturity of jujube fruit. After introducing the improved loss function, P_1, P_2 and P_3 of YOLOv4 have respectively increased by 0.20% to 0.92% compared to the original. The *P* values of CC-YOLO increased by 0.34% to 0.81% compared to the original.

In the case of imbalanced data for fruit classification, there is a noticeable difference in the *P*-values of YOLOv3 and YOLOv4 models for the three classes of fruits. The standard deviations were 0.73 and 0.63, respectively. This indicates substantial differences in the feature extraction capabilities of the models for different fruit classes. After enhancing the loss function, the *P* value of YOLOv4 minimised the disparity in recognition accuracy among categories, reducing the standard deviation to 0.44. The *P* value of CC-YOLOv4 decreased from 0.95 to 0.81 following the improvement of the loss function, thereby reducing the disparity in recognition accuracy among categories. The findings indicate that the enhanced loss function aids in improving the recognition of challenging-to-classify jujube fruit. This helps to minimise the disparity in classification recognition accuracy caused by data imbalance between classes, thereby increasing the overall recognition accuracy.

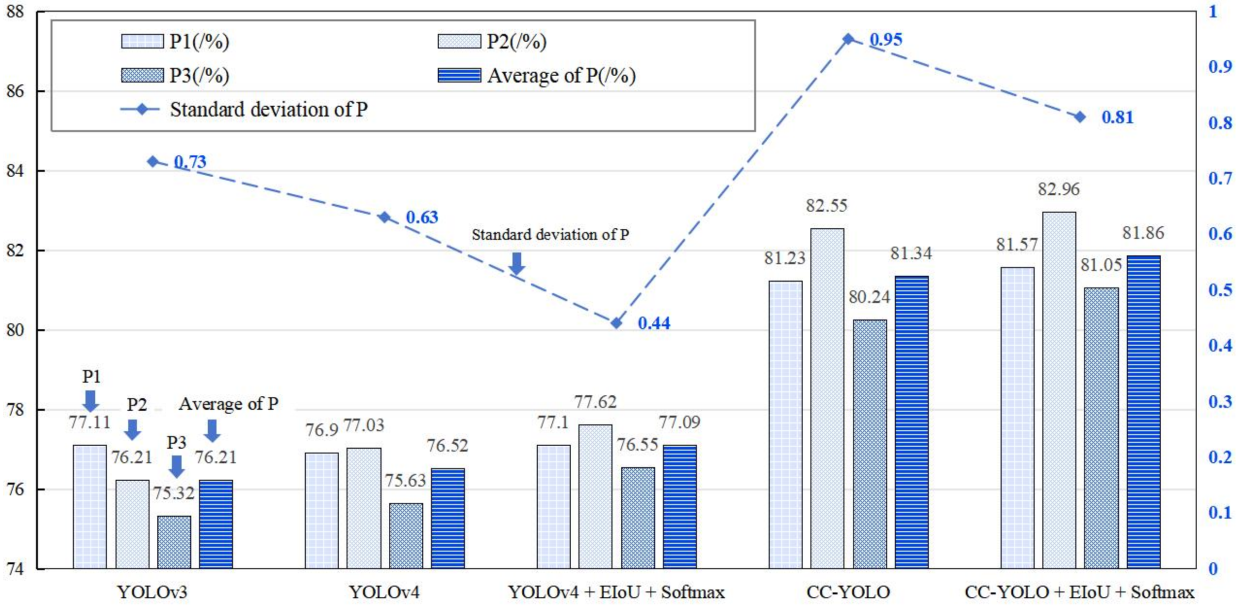


Fig. 5. The *P* values of detection of the models

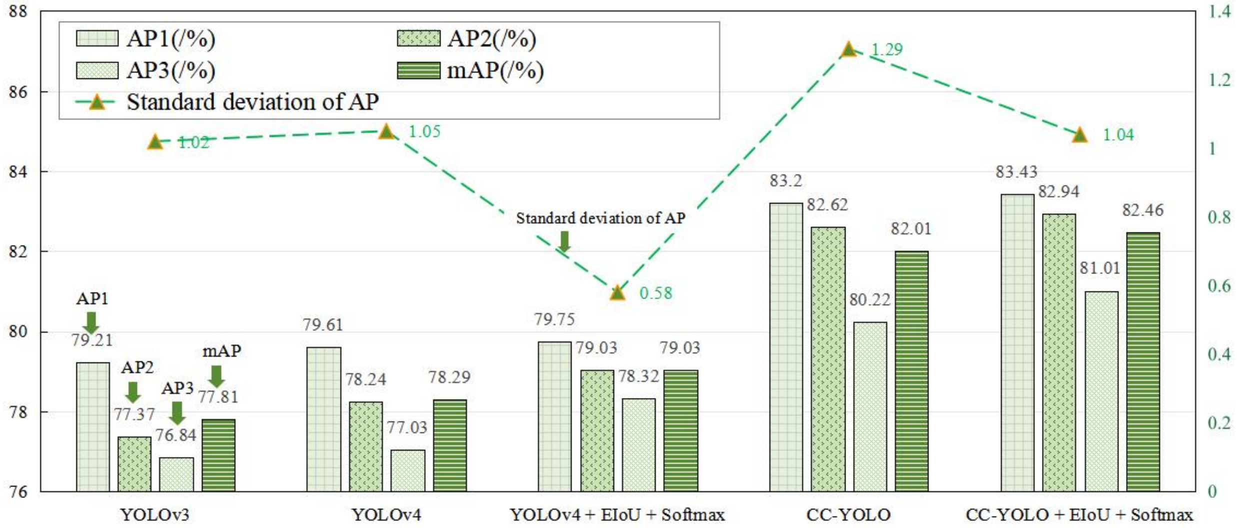


Fig. 6. The *AP* values of detection of the models

In Fig. 6, *mAP* of CC-YOLO, without the improved loss function, is 82.01%, which is 4.20% and 3.72% higher than that of YOLOv3 and YOLOv4, respectively. The AP_1, AP_2, AP_3 of CC-YOLO are 3.38%–5.25% and 3.19%–4.38% higher than those of YOLOv3 and YOLOv4, respectively. AP_2 shows the highest increase, at 5.25% and 4.38% more than YOLOv3 and YOLOv4, respectively.

After the introduction of the enhanced loss function, *mAP* of YOLOv4 increased by 0.74% in comparison to the original model. The enhanced loss function was implemented in CC-YOLO, resulting in an increase of 0.45%. The enhanced loss function demonstrates an improvement in the accuracy of various classes.

The *AP* values for YOLOv3, YOLOv4 and CC-YOLO were significantly different in the case of unbalanced data for the three classes of fruits. Their standard deviations were 1.02, 1.05 and 1.29, respectively. After implementing the enhanced loss function, the variance in *AP* values in YOLOv4 reduced, with the standard deviation decreasing to 0.58. The difference in *AP* values in the model discussed in this paper also decreased, with the standard deviation reducing to 1.04. The enhanced loss function, as demonstrated, is used to diminish the discrepancy in detection accuracy among categories due to data imbalance, resulting in superior classification detection accuracy.

The aforementioned analysis indicates that the enhanced model has significantly increased the accuracy of fruit recognition across various categories compared to the YOLOv3 and YOLOv4 models. Simultaneously, the disparity in detection accuracy across various categories is minimised, mitigating the issue of detection effect discrepancies caused by imbalanced sample data. The enhanced model presented in this paper provides a more accurate recognition of winter jujube fruits across various maturity stages.

2) The accuracy of the predicted boxes

The *IoU* values for the three classes, along with the average *IoU*, are displayed in Fig. 7 for the test set.

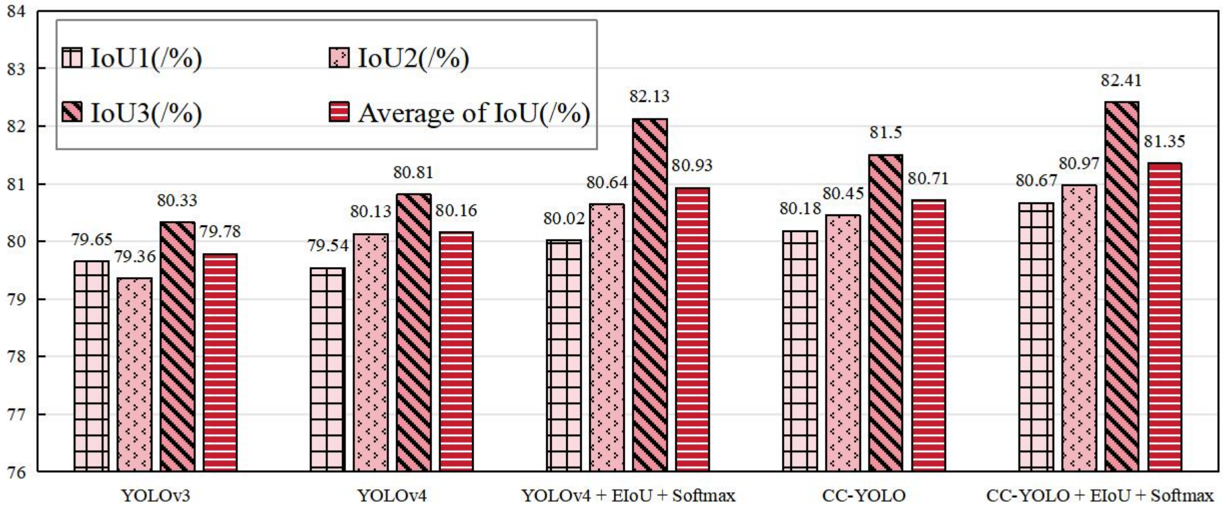
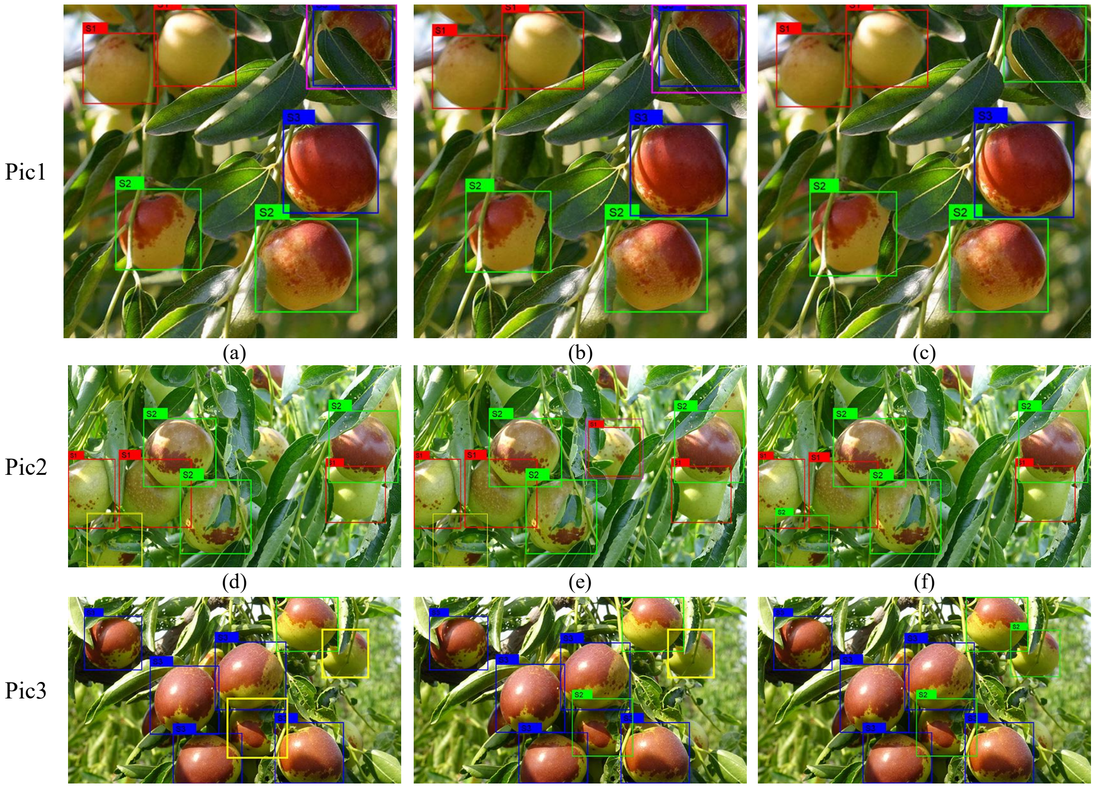


Fig. 7. The *IoU* values of detection of the models



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 8. Detection effects of the models

Given that the *IoU* threshold set during model testing is 75%, the *IoU* values of YOLOv3, YOLOv4 and CC-YOLO are all significantly high. However, the average *IoU* of CC-YOLO is higher, reaching 80.71%, which is 0.93% and 0.55% higher than YOLOv3 and YOLOv4 respectively. *IoU* of the three classes of CC-YOLO improved by 0.53%–1.17% and 0.32%–0.69% compared to YOLOv3 and YOLOv4, respectively. The results indicate that the precision of the predicted boxes' location and size is superior.

After the introduction of the enhanced loss function, the *IoU* values of the YOLOv4 model improved by approximately 0.48% to 1.32% compared to the original model, with an average increase of 0.77%. The *IoU* values of CC-YOLO, with the improved loss function, also increase by 0.49%–0.91% compared to the original model, and the average *IoU* increases by 0.64%. As demonstrated above, the enhanced loss function increases the accuracy of the prediction boxes, thereby improving the precision of target location selection.

3) Detection effect

Three images from the test set, Pic1, Pic2 and Pic3, are selected to compare the detection effects in Fig.8.

The red, green and blue boxes represent the model's prediction boxes for S1, S2 and S3. The yellow box represents an *FN* sample, specifically, a target fruit that should have been detected but was missed. The magenta box represents the *FP* sample, that is, the fruit target detected error class.

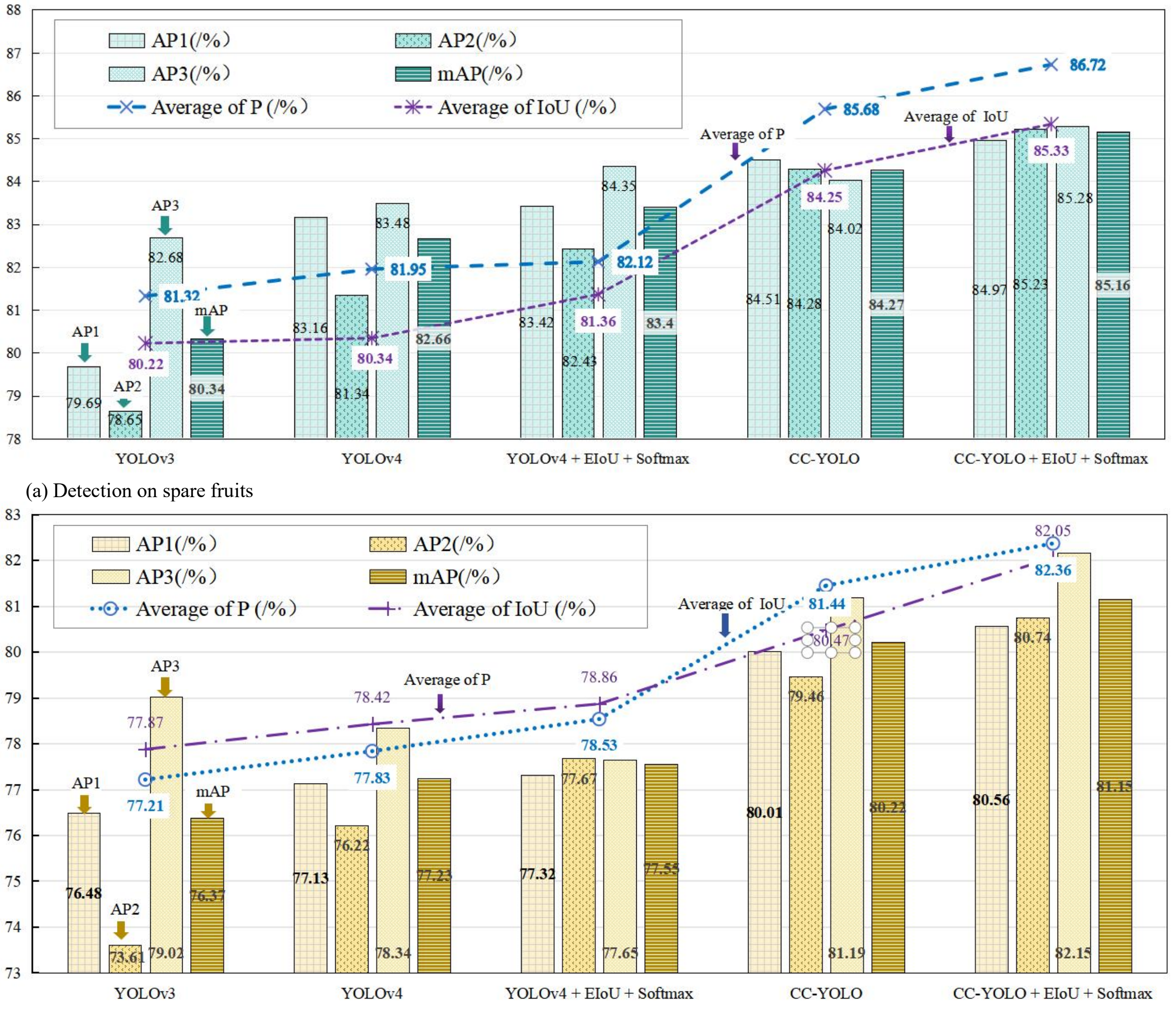
In the figure, (a), (d) and (g) illustrate the detection effects of YOLOv3. (b), (e) and (h) represent YOLOv4, while (c), (f) and (i) represent CC-YOLO. As per Fig. 8, the model presented in this paper demonstrates superior detection accuracy for winter jujube fruit with minimal occlusion in the three images. Conversely, both YOLOv3 and YOLOv4 falsely detected the same position in Image 1, and YOLOv3 missed a detection in Pic2. YOLOv4 demonstrated one false detection and one missed detection in Pic2 and two missed detections in Pic3 in the two models. The likely reason is that the models misclassified a small number of red fruits due to minor occlusion caused by overlapping branches, leaves and fruits, resulting in missed detection. In conclusion, the YOLOv3 and YOLOv4 models exhibit numerous instances of missed and false detections. The proposed model, however, can more effectively circumvent these issues, thereby yielding superior detection results.

F. Detection under different fruit density

Due to the uneven growth distribution of winter jujube fruits, 20 images of different fruit density with clear shooting effect and less background interference were selected separately to form two test sets of of sparse and dense fruits for testing. The evaluation indicators obtained are shown in Fig. 9, and the detection results are shown in Fig. 10.

Compared with the evaluation indicators in the figure, the detection results of the model presented in this paper were significantly higher than YOLOv3 and YOLOv4 in terms of spare and dense fruits. *P*, *mAP*, and *AP* values in the optimal picking stage of S2 were significantly improved by 5.63% and 5.58% than those of YOLOv3 respectively, 2.94% and 3.24% than YOLOv4 respectively. The average *IoU* of this model on the two fruit quantity situation were 84.25% and 80.47% respectively, which were 4.03% and 2.60% higher than YOLOv3, and 3.91% and 2.05% higher than YOLOv4.

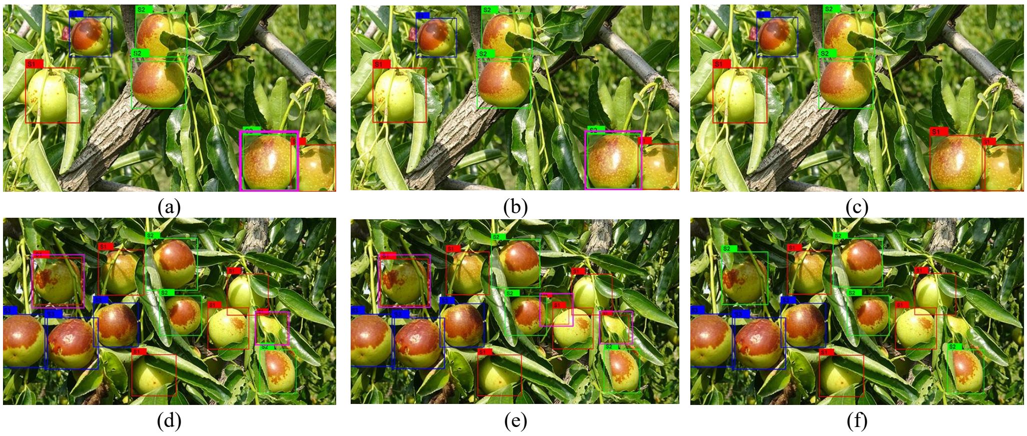
After introducing the improved loss function, both YOLOv4 and CC-YOLO have shown improved detection performance. Specifically, for the model in this paper with the improved loss function, the average of *P* for sparse and dense situations increased by 1.04% and 0.92% respectively compared to before the improvement. The *mAP* values also increased by 0.89% and 0.96%, while the average of *IoU* values increased by 1.08% and 1.58%. The YOLOv4 model also exhibited a slight improvement in detection performance after the introduction of the improved loss



(a) Detection on spare fruits

(b) Detection on dense fruits

Fig. 9. Evaluation indicators of test sets with different densities



(a) (b) (c)

(d) (e) (f)

Fig.10. Detection effects in sparse and dense fruits with the models

function, indicating that the improved loss function can lead to more accurate localization of predicted bounding boxes and higher classification accuracy, thereby enhancing the accuracy of target positioning for harvesting..

It can be observed that there is a significant difference in detection accuracy for the three classes of imbalanced fruits between YOLOv3 and YOLOv4, with the standard deviations of 1.71 and 2.21 for the *AP* values of the three classes, respectively. After using the improved model and loss function proposed, the *AP* values for the three classes of different fruit density has significantly reduced the differences, with the standard deviations reducing to 0.14 and 0.71, indicating the lowest values for the standard deviation. This suggests that the method in this paper has reduced the prominent differences in classification accuracy caused by data imbalances, thus validating the effectiveness of this model under different fruit quantity scenarios.

In Fig.10, (a), (d), (b) and (e) are the detection effects of YOLOv3 and YOLOv4 of sparse and dense fruits under the same images, while (c) and (f) are those of CC-YOLOv4. In (a) and (b), the S1 class of fruit is mistakenly detected as the S2 class. This may be due to the fact that the fruit in the picture is slightly red caused by sun exposure, rather than the deep red color that represents maturity. In (d) and (e), the fruit of S2 class was mistakenly detected as S1 class at the same position. This may be due to the influence of light, shadow and the slight occlusion of branches and leaves, affecting the detection of the red parts of winter jujube fruits. In addition, most of the occluded fruit caused by overlapping fruit was misdetected in (e). The proposed method in this paper avoided the above situation.

It can be seen that the proposed method has a higher detection accuracy for the classification of fruit maturity in different density. The method is more accurate in the detection of more complete, easy-to-pick fruits. Therefore, the method proposed in this paper is relatively robust and meets the recognition requirements in orchards.

V. CONCLUSION

In this paper, a detection method was proposed for detecting winter jujube fruits in orchards. The experimental results indicate that the *P*, *mAP* and *IoU* values for the three classes of winter jujube fruits were 81.86%, 82.46% and 81.35%, respectively. There is a significant reduction in the overall number of model parameters, with the count standing at only 26.9M. The model size is 108MB, a comparison that highlights its compactness relative to the YOLOv3 and YOLOv4 models. In batch detection, images with higher resolution can achieve 28.8 FPS, satisfying the real-time identification needs for the automatic selection of winter jujubes.

In practical applications, the impact of light on the recognition effect should also be considered. It is necessary to conduct further experiments under various weather conditions in the actual light environment to enhance the classification and recognition accuracy while ensuring real-time performance.

REFERENCES

[1] L. Ma, Research on the Countermeasures for the Development of the Modern Jujube Industry in Hebei Province. Baoding: Hebei Agricultural University, 2012.

[2] J. C. Duan, Research on the Characteristic Agricultural Disaster Risk Guarantee System in Cangzhou City, Hebei Province: The Case of Winter Jujube. Kunming: Yunnan University of Finance and Economics, 2016.

[3]M. Y. LI, G. Z. SHi, J. R. Z. Hu et al. Comprehensive evaluation of fruit quality of fresh jujube winter jujube. Non-Wood Forest Res., vol. 39, no. 4, 2021, pp. 256–263.

[4] Y. T. Wang, J. M. Kan, W. B. Li, and C. Zhan, 'Image segmentation and maturity recognition algorithm based on color features of Lingwu long jujub', Adv. J. Food Sci. Technol., vol. 5, no. 12, Dec. 2013, pp. 1625–1631.

[5] S. Bi, F. Gao, J. W. Chen, and L. Zhang, 'Detection method of citrus based on deep convolution neural network', Trans. CSAM, vol. 50, no. 5, May. 2019, pp. 181–186.

[6] Y. Liu, L. Y. Ye, X. D. Sun, H. Ru-bing, X, Huai-chuan, M. Kui-rong, Z. Dan-ning, and W. Ming-ming, 'Maturity evaluation model of tangerine based on spectral index', Chin. Opt., vol. 11, no. 01, Feb. 2018, pp. 83–91.

[7] J. Zhao, P. K. Quan, M. J. Ma, L. Li, D. He, and H. Zhang, 'Comparative analysis of harvest maturity model for Fuji apple based on visible/near spectral nondestructive detection', Trans. Chin. Soc. Agric. Mach., vol. 49, no. 12, Dec. 2018, pp. 347–354.

[8] Z. H. Lu, M. F. Zhao, J. Luo, G. Wang, and D. Wang, 'Design of a winter-jujube grading robot based on machine vision', Comput. Electron. Agric., vol. 186, no. 5, Jul. 2021, p. 106170.

[9] T. W. Wang, Y. Zhao, Y. X. Sun, R. Yang, Z. Han, and J. Li, 'Recognition approach based on data-balanced faster R-CNN for winter jujube with different levels of maturity', Trans. Chin. Soc. Agric. Mach., vol. 51(suppl. 1), 2020, pp. 457–463 + 492.

[10] T. Z. Liu, G. F. Teng, Y. C. Yuan, et al. 'Fruit recognition method of winter jujube in natural scene based on improved YOLOv3', Trans. Chin. Soc. Agric. Mach., vol. 52, no. 5, 2021, pp. 17–25. (in Chinese with English abstract)

[11] H. Zhao, Y. J. Qiao, H. J. Wang, and Y. Yue, 'Apple fruit recognition in cChin. Soc. Agric. Eng., vol. 37, no. 16, 2021, pp. 127–135.

[12] D. A. Zhao, R. D. Wu, X. Y. Liu, and Y. Zhao, 'Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background', Trans. Chin. Soc. Agric. Eng., vol. 35, no. 3,

2019, pp. 164–173.
[13] Z. P. Cao, R. B. Yuan, ‘Real-time detection of mango based on improved YOLOv4’, Electronics, vol. 11, no. 23, Nov. 2022, p. 3853.
[14] L. S. Wang, M. X. Qin, J. Y. Lei, X. Wang, and K. Tan, ‘Blueberry maturity recognition method based on improved YOLOv4-Tiny’, Trans. Chin. Soc. Agric. Eng., vol. 37, no. 18, 2021, pp. 170–178.
[15] T. B. Huang, H. Q. Huang, Z. Li, S. Lu, X. Xu, Q. Dai, and W Wen. ‘Citrus fruit recognition method based on the improved model of YOLO v5’, J. Huazhong Agric. Univ., vol. 41, no. 4, Oct. 2022, pp. 170–177.
[16] H. X. Sun, S. J. Zhang, R. Ren, L. Y. Su, ‘Surface defect detection of ‘Yuluxiang’ pear using convolutional neural network with class-balance loss’, Agronomy, vol. 12, no. 9, Aug. 2022, p. 2076.
[17] Z. Y. Zhang, M. Y. Luo, S. X. Guo, G. Liu, S. Li, and Y. Zhang, ‘Cherry fruit detection method in natural scene based on improved YOLO v5’, Trans. Chin. Soc. Agric. Mach., vol. 53(suppl. 1), 2022, pp. 232–240.
[18] A. Bochkovskiy, C. Y. Wang, and H. Y. Mark Liao, ‘YOLOv4: Optimal speed and accuracy of object detectio’, Comput. Electron. Agric., vol. 186, no. 5, Apr. 2021, p. 106170.
[19] D. H. Wu, S. C. Lv, M. Jiang, and H. Song, ‘Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments’, Comput. Electron. Agric., vol. 178, Nov. 2020, p. 105742.
[20] Z. Yu, Y. Shen, C. Shen, ‘A real-time detection approach for bridge cracks based on YOLOv4-FPM’ Autom. Constr., vol. 122, Feb. 2021, p. 103514.
[21] C. Y. Wang, H. Y. Liao, Y. H. Wu, P.-Y. Chen, J.-W. Hsieh, and I-H. Yeh, ‘CSPNet: A new backbone that can enhance learning capability of CNN’, Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 390–391.
[22] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, ‘CBAM: Convolutional block attention module’, Proc. European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[23] Y. F. Zhang, W. Q. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, ‘Focal and efficient IOU loss for accurate bounding box regression’, Neurocomputing, vol. 506, Sep. 2022, pp. 146–157.