Research on High Performance Apple Recognition Model Construction Strategy Based on Deep Learning

Yixuan Wang *, Zhengkun Su

School of Communication, Qufu Normal University, Rizhao, China, 276826

* Corresponding Author Email: 2956131489@qfnu.edu.cn

Abstract. In order to improve the working efficiency and environmental adaptability of apple picking robots, which can realize accurate recognition and localization, ripening assessment, and quality assessment under complex situations such as "leaf occlusion", "branch occlusion", "fruit occlusion", "mixed occlusion", etc., this paper proposes a Faster R-CNN-based apple recognition method. "mixed occlusion" and other complex situations to achieve accurate recognition and localization, maturity assessment, quality assessment, the paper proposes an apple recognition method based on Faster R-CNN. The method combines the cv2.inRange and cv2.fitEllipse functions in OpenCV to optimize the model, and performs color segmentation and shape fitting by setting up color and shape thresholds to recognize apples. The trained model was tested to have a recall of 88.44% under the validation set, an accuracy of 90.35%, an F1 value of 89.38%, and a recognition time of about 0.3 s per image.The experimental results showed that the number of apples in the dataset showed a normal distribution, with the number distributed more in the range of 0 to 10. Candidate frames are screened by RPN, and apple coordinates are derived based on the coordinate transformation formula, and the experimental results show that the apples are more densely distributed near the range of x=100, y=100. The image color distribution was extracted through color histogram, and the entropy value of GLCM, contrast as the texture feature of apples, and the classification prediction of apple ripeness was carried out by using Random Forest Model, and the accuracy of the model prediction was 83.2%, the precision rate was 84.1%, and the recall rate was 83.5%, and the experimental results showed that 88.7% of the apples were ripe. Image edges were extracted by the Canny edge detection algorithm and the area values of pixel points within the edges were calculated. The ratio of pixel area converted to apple mass was estimated to be 0.007 based on the image resolution and apple density, and the test results indicated that the mass of apples was mostly concentrated in the range of 0~100g.

Keywords: Computer Vision, Apple Recognition, Faster R-CNN, Deep Learning.

1. Introduction

Currently domestic and international research on apple recognition has made some progress. Color, shape, and texture are important features used by picking robots to distinguish between fruit and background images[1] , LIU et al.[2] used an SVM classifier based on color and shape features to detect apples in real time with an accuracy of 90.6%. However, the recognition accuracy is limited by the quality of the target's feature information, which is not applicable to the real-time recognition and detection of fruits in the orchard in complex situations. For long fruits, Meng et al.[3] use edge detection algorithm to extract the boundary line between the target and the image background to detect the target, and use Canny operator and Hough transform to detect the target contour so as to realize the accurate recognition and localization of fruits.

In recent years, target detection algorithms based on deep learning have been gradually applied in the field of fruit recognition, Abdellah El Zaar et al.[4] utilized deep convolutional neural networks based on the modified fine-tuning method (MFTs-Net) for date recognition, and the experimental results showed that the model can recognize different date categories of jujubes with a high accuracy rate. K.Y. Ren et al.[5] utilized the YOLOX model, optimized the decoupled detection head, and proposed a new method combining Mosaic and MixUp in data enhancement, which is a great improvement over the related algorithms such as YOLOv5, and achieved an accuracy of 98.6% for fruit recognition.Gill Harmandeep Singh et al.[6] proposed the Type - II Fuzzy, TLBO (Teacher-

Learner Based Optimization) and the application of Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM) in Deep Learning for enhancement, segmentation, recognition and classification of fruit images with excellent results.WAN et al.[7] Improved the pooling and convolutional layer structure of the existing Faster R-CNN model and designed a multi-category model for fruit recognition. improved and designed a multi-category fruit detection method with an average detection accuracy of 91%. In addition, some scholars add attention mechanisms such as CA[8] , SE[9] , CBAM[10] to the convolutional layer of the detection model to further improve the robustness and generalization of the model in the complex orchard environment.

In order to realize accurate and efficient recognition of multiple information of apples (number, location, ripeness, quality) in complex orchard situations, this paper designs an apple recognition method based on Faster R-CNN. The method combines the cv2.inRange and cv2.fitEllipse functions in OpenCV for color segmentation and shape fitting to identify and locate apples. Random forest model based on dual features of color and texture is also used for classification prediction of apple ripeness, and image edges are extracted by Canny edge detection algorithm, and area values of pixels within the edges are calculated to estimate the quality of apples. (Data Sources: http://www.apmcm.org/).

2. Dataset processing

To improve the efficiency of the model, the image is normalized, grayscaled, and Gaussian filtered before constructing the model.

Image normalization adjusts the pixel values of an image so that they fall into a specific range, usually the pixel values are normalized to a specific range, such as [0, 1] or [-1, 1], in order to better fit the training of the model. In this paper, Min-Max is used for normalization with the following formula:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

Where x is the original data and is the normalized data, and and represent the maximum and minimum values in the original data, respectively.

Image Grayscaling is the process of converting a color image into a grayscale image with the following formula:

$$gray = R * 0.299 + G * 0.587 + B * 0.114 \tag{2}$$

Where. R, G, B denote the pixel values of the red, green, and blue channels of the image, respectively. Apple grayscale is shown in Fig 1, Fig 2:

Fig 1. Original image of apple **Fig 2.** Image of apple after greyscaling

Gaussian filtering is based on the properties of the Gaussian function, where the center pixel has the highest weight, the surrounding pixels have decreasing weights, and the correlation between pixels is determined by distance. The pixels are weighted and averaged to reduce the component of high frequency details in the image to blur the image and remove noise. The formula is given below:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

Where x, y are the positions of the pixels in the filter, respectively, and is the standard deviation.

3. Apple recognition and counting model based on Faster R-CNN

3.1. The establishment of Faster R-CNN

The essential features of apples that distinguish them from leaves and branches are their characteristic red (or cyan) color and their oval shape that is close to a circle. Therefore, in this paper, feature extraction is carried out from the two dimensions of color and shape.

In this paper, we use the cv2.inRange function in the OpenCV library to implement color segmentation. The HSV value of red color is set from [0, 40, 40] to [10, 255, 255] and the HSV value of cyan color is set from [35, 40, 40] to [85, 255, 255]. The shape of the apple approximates an ellipse, and in this paper, we utilize the cv2.fitEllipse function in the OpenCV library for ellipse fitting to find an ellipse that best fits the contour points by least-squares fitting to those points. After the ellipse is fitted, a eccentricity threshold (0.5 in this paper) is set to filter out redundant shapes. An area threshold (50 for this paper) is also set to filter out ellipses that are recognized as too small (these ellipses are most likely not apples).

According to the features of apple, this paper chooses Faster R-CNN model to recognize apple image. The input image, after convolution, activation, pooling and other steps, gets the feature map, the feature map enters the RPN network to get the suggestion box, and then enters the ROI Pooling layer with the previous feature map for feature fusion to get the suggested feature map, and finally enters the classification and regression network, which performs the classification through the Softmax function as well as the use of the bounding box regression to get the positional offset bbox_ pred, which is used to obtain a more accurate detection box. In the input layer image is reshaped to M×N size and enters the convolutional layer undergoes 15 convolutional operations, 15 activation functions to increase the nonlinearity of the convolutional network. In RPN layer feature maps undergo convolution operation to get suggestion frames into ROI Pooling layer.ROI Pooling layer collects the suggestion frames obtained through RPN layer and fuses the suggestion frames with the feature maps obtained through convolution layer to get the suggested feature maps. After entering the classification layer, the suggestion feature map passes through the fully connected layer, and then through a series of operations such as the Softmax function, the category of each suggestion box and the probability of the category to which the suggestion box belongs are calculated, and at the same time, the bounding box regression is used again to get the positional offsets, which makes the accuracy of the labeled boxes more accurate.

In this study, Cross Entropy Loss is used for classification loss function and Smooth L1 Loss is used for regression loss function. Cross Entropy Loss is used to measure the difference between two probability distributions. Cross Entropy Loss formula is as follows:

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log\left(q(x_i)\right) \quad (4)$$

Where p is the true probability distribution and q is the predicted probability distribution of the model. Smooth L1 Loss is used for the prediction of object bounding box position in target detection with the following equation:

$$Soomth_{L1}(y,\bar{y}) = \frac{1}{N}\sum_{i=1}^{N} Soomth_{L1}(y,\bar{y}) \quad (5)$$

y is the actual value, $\bar{y}$ is the predicted value of the model, and N represents the number of samples.SGD optimization algorithm is set to the function to train the machine learning model thus minimizing the loss function. The image its preprocessed and labeled and then the dataset is fed into the model for model pre-training.The Faster R-CNN architecture is shown in Fig 3:

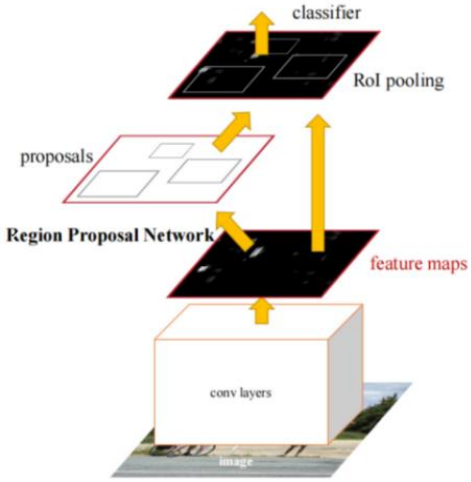

Fig 3. Schematic of Faster R-CNN architecture

The experimental results show that the recall of the model is 88.44%, the accuracy is 90.35%, the F1 value is 89.38%, and the recognition time of each image is about 0.3s, which indicates that the model has high recognition accuracy and image processing speed.

3.2. Apple recognition and counting

The pre-trained Faster R-CNN model image is called for recognition and the recognition effect is shown in Fig 4:

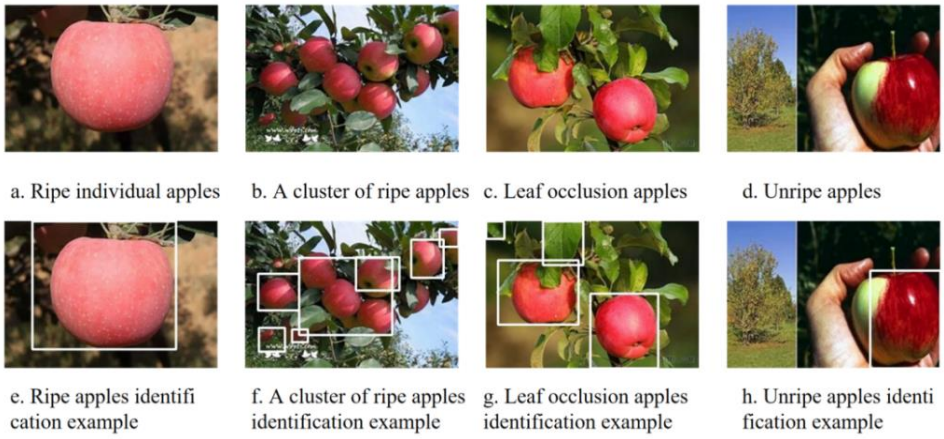

Fig 4. Graph of apple recognition effect under different conditions

The number of apples identified was counted and a histogram of the distribution of the number of apples was plotted as in Fig 5:

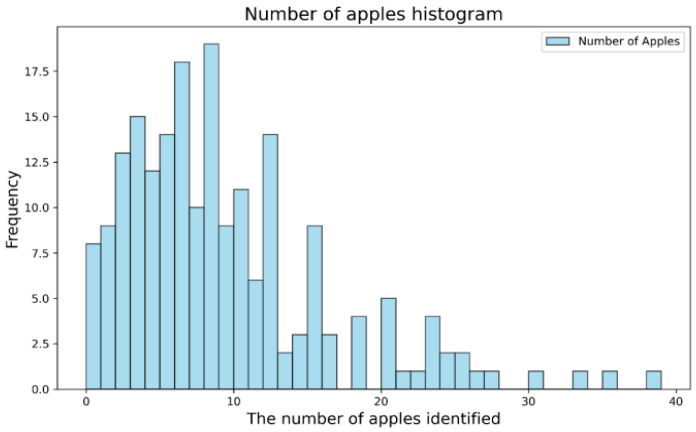

Fig 5. Number of apples histogram

According to the histogram, the distribution of the number of apples is generally normally distributed, with slightly more distributed in the range of 0 to 10.

4. Apple location calculation model based on Faster R-CNN

Faster R-CNN achieves accurate bounding box extraction of targets in images through RPN with bounding box regressors. In the candidate box generation stage, the input image is subjected to feature extraction through a shared convolutional basis, followed by the introduction of a sliding window mechanism on the feature map, and the model generates a series of candidate boxes. In the bounding box regressors stage, the candidate boxes screened by the RPN are passed as regions of interest to the next stage.RoIs are mapped into fixed-size features by RoI Pooling or RoI Align, and subsequently processed through a series of fully connected layers. The model performs two key tasks simultaneously: predicting the class of targets within the RoI; and adjusting the RoI bounding box. The output of this stage is the bounding box coordinates for each target.

The coordinates of the bounding box are usually expressed in the upper left and lower right corners. Each bounding box coordinate is $(x_{min}, y_{min}, x_{max}, y_{max})$, where (x_{min}, y_{min}) is the coordinate of the upper left corner, where (x_{max}, y_{max}) is the coordinate of the lower right corner. In this paper, the lower left corner is the origin and is calculated as follows:

$$width = x_{max} - x_{min}, height = y_{max} - y_{min} \tag{6}$$

$$x' = \frac{width}{2}, y' = \frac{height}{2} \tag{7}$$

The orientation of each apple is derived from the formula and the coordinates of each apple were plotted on a scatter plot as in Fig 6:

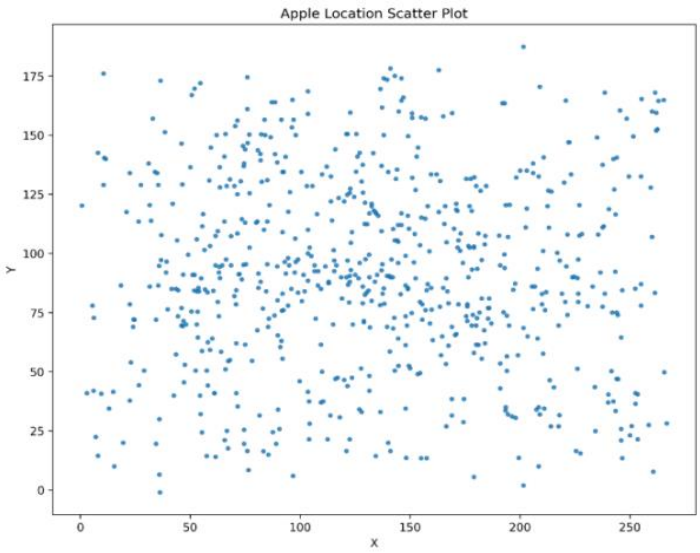

Fig 6. Apple Location Scatter Plot

Observation of the image shows that apples are more densely distributed around the range x=100, y=100.

5. Random forest predicts apple ripeness

This study is oriented to the ripeness assessment of apples, and the external manifestation of ripeness is the color and texture of apples, so the ripeness characteristics of apples are extracted from the two dimensions of the color and texture of apples. The cv2.cvtColor function and cv2.calcHist function in OpenCV can be used to complete the color histogram. An example of color histogram of apple image is shown in Fig 7:

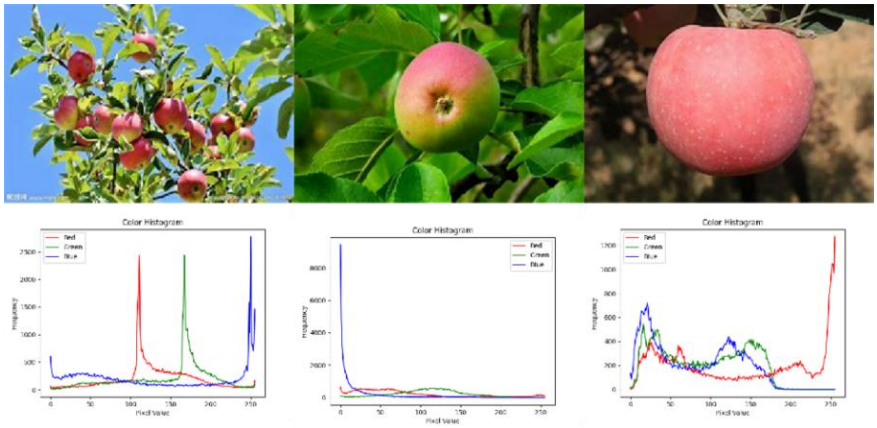

Fig 7. Apple image colour histogram

It is not enough to identify the ripeness of apples by color recognition, but also need to consider the texture information of apples. In this paper, the entropy and contrast of GLCM are chosen as texture features. Its formula is as follows:

$$Entropy = -\Sigma_i \Sigma_i C_{ij} log C_{ij}, Contrast = \Sigma_i \Sigma_i (\mathrm{i} - j)^2 C_{ij} \quad (8)$$

Where C_{ij} is a pixel point C with horizontal coordinate i and vertical coordinate j. The apple image GLCM visualisation is shown in Fig 8:

Fig 8. GLCM visualisation

Random forest model is selected for prediction. The number of decision trees is initially selected as 100, the random seed tree is 42, and the depth of the decision tree is 10, and the dataset is prepared to train the random forest model. Comparing the prediction results with the true values, the accuracy of the model prediction is 83.2%, the precision rate is 84.1%, and the recall rate is 83.5%, which indicates that the model has a good capability of maturity assessment.

The resulting histogram of apple ripeness is shown in Fig 9:

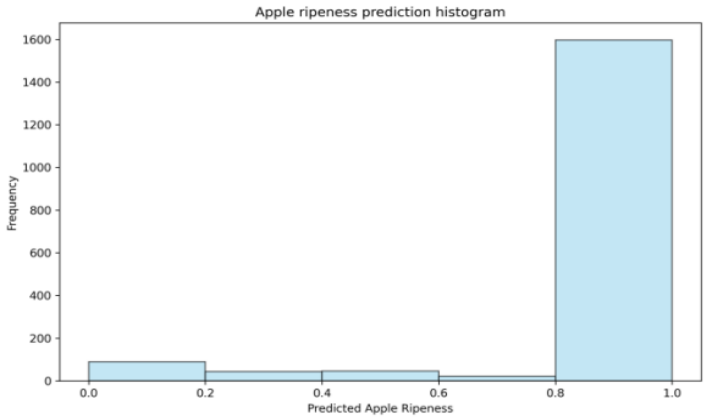

Fig 9. Apple ripeness prediction histogram

Observation of the histogram shows that the vast majority of the apples are ripe, calculating that about 88.7% of the apples are ripe.

6. Apple quality measurement model based on pixel area transformation

6.1. Pixel extraction

Apple edges are extracted using Canny edge detection algorithm and pixel points are extracted within the edges. The image gradient magnitude, gradient direction is calculated using the following formula:

$$G = \sqrt{G_x^2 + G_y^2}(\text{math.})\ \text{genus}\theta = tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (9)$$

G_x is the horizontal gradient, G_y is the vertical gradient. Water is vertical square.Apple Canny edge extraction is shown in Fig 10:

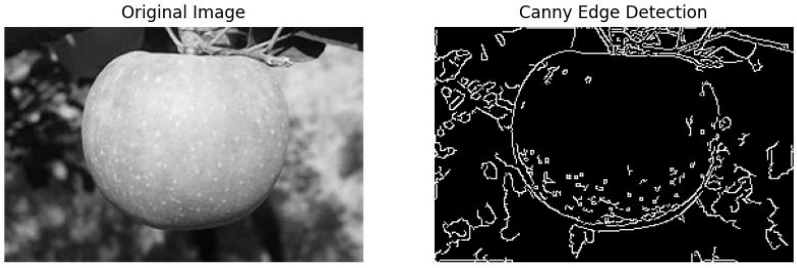

Fig 10. Example of Canny edge extraction

The pixel point extraction formula is as follows:

$$s_p = \sum_{x,y} I(x,y) \quad (10)$$

$s_p$is the area of all pixel points and (x, y) are the coordinates of each pixel point. An area threshold is set up to filter clusters of pixels that are too small (to prevent possible problems with recognition as leaves).

6.2. Area-mass conversion

The resolution of the image is 270*185, so an image can be considered to have 49,950 pixel points, the actual size of the image on the screen is about 2.7cm*1.85cm, and the density of the apple is about 0.9g/cm^3 . Based on the above information it can be estimated that the conversion ratio of apple area to mass is 0.007, i.e. a pixel point can be considered to weigh 0.007 g. A histogram of the mass distribution of all the apples is plotted as in Fig 11:

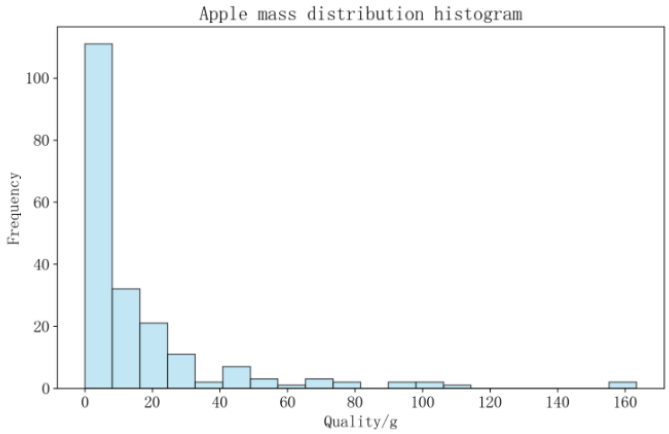

Fig 11. Apple mass distribution histogram

Observation of the images shows that the mass of the apples is mostly concentrated between 0 and 100g.

7. Conclusions

1) In this paper, Faster R-CNN is chosen as the algorithm for recognizing fruits, and the algorithm can complete the apple recognition work well. At the same time, in solving the problems of branch and leaf occlusion as well as the interference of small fruit recognition, this paper proposes that the color segmentation and shape fitting can be combined with the cv2.inRange and cv2.fitEllipse functions in OpenCV, combined with the SGD optimization algorithm, to extract the features of the apples, and the accuracy of the recognition of the apples can be effectively improved. The experimental results show that the recall rate is 88.44%, the accuracy rate is 90.35%, the F1 value is 89.38%, and the recognition time is about 0.3s per image.

2) Based on the improved Faster R-CNN model for recognition and localization of apple images, the experimental results show that the distribution of the number of apples is roughly normal, more distributed in the range of 0 to 10, and apples are more densely distributed in the range of x=100, y=100.

3) In this paper, we consider the less studied problems of ripeness prediction and quality estimation in the field of fruit recognition research, using random forest for the prediction of ripeness of apple fruits, and a creative and simple method, pixel to area conversion, for the estimation of the quality of apples, and the accuracy of the random forest model prediction is 83.2%, the precision rate is 84.1% and recall rate is 83.5%. The experimental results show that about 88.7% of the apples in the dataset are ripe, and the quality of the apples is mostly small, mostly concentrated in the range of 0~100g.

References

[1] Taixiong Zheng, Mingzhe Jiang, Mingchi Feng. A review of research on vision-based target recognition and localisation methods for picking robots[J]. Journal of Instrumentation,2021,42(09):28-51.

[2] LIU X, ZHAO D, JI W, et al. A detection method for apple fruits based on colour and shape features[J]. IEEE Access, 2019, 7: 67923-67933.

[3] MENG Y H, WANG J Q, TIAN E L, et al. Research on navigation of agricultural UAV based on single chip microcomputer control [J]. Mechanization Research,2020, 42(3): 245-248.

[4] Zaar* E A, Assawab R,Aoulalay A, et al.MFTs-Net: A Deep Learning Approach for High Similarity Date Fruit Recognition[J].Journal of Advances in Information Technology,2023,14(6).

[5] Keying R, Xiaoyan C, Zichen W, et al. Fruit Recognition Based on YOLOX*[J]. Proceedings of the International Conference on Artificial Life and Robotics,2022,27470-473.

[6] Singh H G, Ganpathy M, Singh B K, et al.Fruit recognition from images using deep learning applications[J].Multimedia Tools and Applications,2022, 81(23):33269-33290.

[7] WAN Shaohua, GOUDOS S. Faster R-CNN for multi-class fruit detection using a robotic vision system[J]. Computer Networks, 2019, 168: 107036.

[8] MA J, LU A, CHEN C, et al. YOLOv5-lotus an efficient object detection method for lotus seedpod in a natural environment[J]. Computers and Electronics in Agriculture, 2023, 206: 107635.

[9] HU Weixin, XIONG Juntao, Liang Junhao, et al. A method of citrus epidermis defects detection based on an improved YOLOv5[J].14 Biosystems Engineering, 2023, 227: 19-35.

[10] CAO Z, MEI F, ZHANG D, et al. Recognition and detection of persimmon in a natural environment based on an improved YOLOv5 model[J]. Electronics, 2023, 12(4): 785.