Extraction and Recognition of Robotic Apple Picking Image Features Based on YOLOv5 Detection Models

Wenqian Hong[1], Hao Wu[2], and Yirong Jiang[2*]

[1]School of Mathematics and Statistics, Guilin University of Technology, Guilin, Guangxi 541004, P.R. China

[2]School of Mathematics and Sciences, Guangxi Minzu University, Nanning, Guangxi, 530006, P.R.China

*Corresponding author: 20240021@gxmzu.edu.cn

Abstract

As China has emerged as one of the leading exporters of apples globally, the shortage of agricultural labor has posed a significant challenge for the apple industry's growth. To address the issue of image recognition for robotic apple picking in complex orchard settings, this paper combines computerized image processing and deep learning concepts to propose a detection model based on YOLOv5. By implementing image preprocessing techniques and optimizing the loss function, the study successfully achieves accurate extraction and recognition of apple image features. The experimental findings demonstrate the high performance and accuracy of the proposed method in apple picking tasks, offering valuable support for the advancement of robotic automated picking systems. Future research endeavors will focus on further refining the algorithm to enhance efficiency in real-world production settings. The improved OLOv5 Detection Models proposed in this article can be applied in fields such as industrial detection, intelligent traffic signal control, and sports events.

Index Terms—YOLOv5 detection model, digital image processing, apple feature extraction and recognition, labelImg

1 Introduction

China, as the world's largest exporter of apples, has most local farmers growing apples in their own orchards. During the apple picking season, a significant number of workers are required to harvest ripe apples. The rapid urbanization in China, along with an aging agricultural workforce and a large portion of young individuals seeking work opportunities elsewhere, has resulted in labor shortages during this crucial season[1]. To address this issue, China has been developing apple-picking robots since 2011, making notable advancements. However, existing robots often struggle to accurately identify various obstacles in the orchard environment, such as 'leaf shading', 'branch shading', 'fruit shading', and 'mixed shading'. Without precise judgment based on the actual conditions, harvesting directly can lead to significant damage to the fruits, as well as potential harm to the robotic arm and the workers. This can negatively impact both harvesting efficiency and fruit quality, resulting in substantial losses. Furthermore, the accurate identification and classification of harvested fruits is crucial for subsequent sorting, processing, packaging, and transportation. Moreover, the similarity in color, shape, and size between apples and other fruits poses a challenge for apple-picking robots in accurately distinguishing between them.

At present, many scholars have studied this problem. Relevant studies mainly include:Wang Dandan etal[2] further analyzed the problems in the vision system of apple picking robot to provide reference for the in-depth study of the vision system of apple picking robot.Ka-pach etal[3] investigated the apple color detection method, but the algorithm detection effect is not ideal for immature apples or the situation of having branches and leaves blocking, and apples are similar to the background, and so on. Cao Chunqing etal [4] realized accurate recognition and 3D localization of apples in multiple natural scenes by fusing YOLOv3 and binocular vision algorithms. Zhao De'an etal [5] proposed a YOLO deep convolutional neural network-based localization method for robotic apple picking in complex backgrounds, using optimized YOLOv3 deep convolutional neural network to locate apples, and achieved apple recognition and localization in complex environments. Cao Zhipeng etal [6] used YOLOv4 neural network can recognize apples better, but the recognition speed of YOLOv4 is low, which can't meet the demand of real-time picking. The above methods perform well in recognizing apple targets, but they require high computational resources.

From the above research, we found that these methods are difficult to achieve fast and accurate identification and localization. It will be interesting to study what happens to a method if it not only enables quick recognition but also ensures precise localization. The proposed approach involves developing an apple image recognition model using the depth-separable convolutional YOLOv5 model, with optimized loss functions to enhance speed and accuracy in recognizing apples in complex environments. This advancement further facilitates the practical implementation of domestic apple picking robots. By analyzing labeled apple images and extracting relevant features, a high recognition rate, speed, and accuracy are achieved. Moreover, data analysis of the images enables automatic calculation of the number, location, maturity, and estimated quality of apples, thereby improving fruit recognition rates. The results obtained exhibit high precision and recall

rates of 99.08% and 99.3%, respectively. This research holds significant implications for the advancement of robotic apple picking technology in China.

2 Image Preprocessing

2.1 Description of the experimental dataset

To ensure the complexity of the apple image data, this paper uses the dataset of the 13th APMCM Asia-Pacific Regional Undergraduate Mathematical Modeling Competition problemA, 2023.There are three subsets of this dataset and they are: subset 1, subset 2 and subset 3. The basics are as follows:

Subset 1 is a ripe apple image dataset containing 200 images of ripe apples, each with a size of 270 * 180 pixels. Some screenshots of Subset 1 are shown in Fig.1:

Figure 1: Mature apple image datasets

Subset 2 is a fruit picking image dataset containing 20705 images of different picking fruits, each with known labels and classifications, with a size of 270 * 180 pixels. Some of the screenshots in Subset 2 are as requested in Fig.2, which includes apples, cactus fruits, pears, plums and tomatoes.

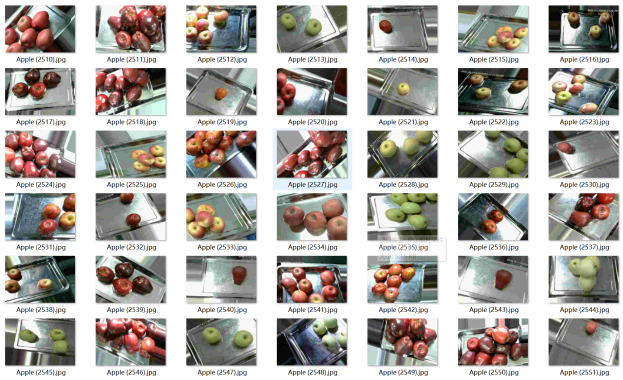

Figure 2: Fruit picking image dataset

Subset 3 shows the labeled dataset containing 20705 images of different picked fruits, each with a size of 270 * 180 pixels, but with unknown labels and classifications. Some screenshots of Subset 3 are shown in Fig.3:

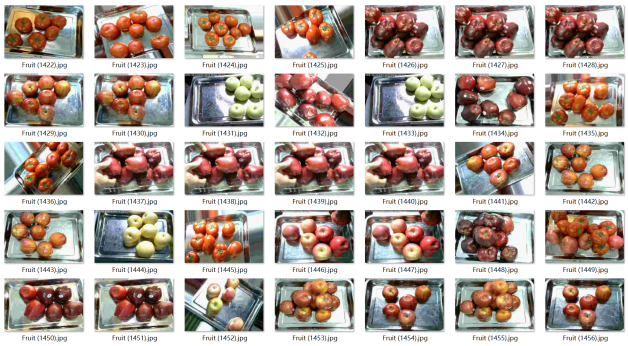

Figure 3: Tagged data sets

2.2 Original Image Information

In the apple images given in Subset 1, there are a total of 200 images, all of which have pixels of 270 x 185. The image file was taken at basically the same time and with sufficient light. However, the images are divided into four parts; a portion of the ripe red apples have problems such as part of the image being blurred, overexposed or underexposed, and shadows being blocked by leafy branches; a portion of the apples have people blocking them; a portion of the images are images of immature green apples or blossom bones; and a portion of the images are of other fruits. Some of the images are shown in Fig.4:

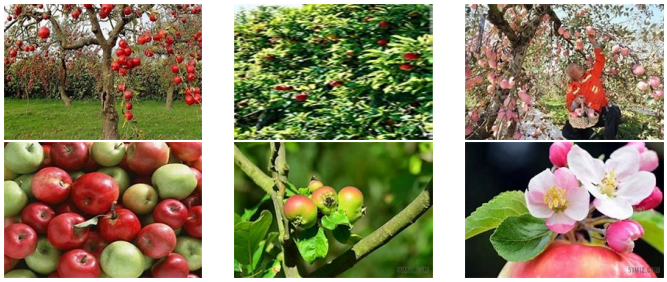

Figure 4: Example of an image from part of the subset

The above types of images will affect the feature extraction of apples, so it is necessary to pre-process the apple image in Subset 1 with image denoising, image enhancement, color space conversion, etc. to enhance the contrast of the image.

2.3 Image preprocessing methods

The steps of image preprocessing are shown in Fig.5:

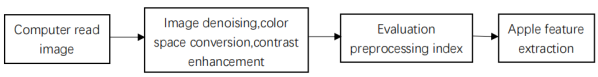

Figure 5: Steps in image preprocessing

Among them, we use median filtering method, inverse sharpening mask method of image denoising, enhancement, to

avoid the image blurring and other problems to interfere with the later for the color, shape and other features of the apple extraction.

2.4 Rating system

In order to judge the advantages and disadvantages of the preprocessed images, standard deviation standrad and structural similarity SSIM are selected as the evaluation indexes of the preprocessing results. SSIM index is mainly from the brightness, contrast and structure of three aspects to measure the degree of similarity between the two images, the value range is [0,1]. The SSIM index mainly measures the degree of similarity between two images from three aspects: brightness, contrast and structure, and the value range is [0,1], the larger the SSIM value is, the more similar the structure of the two images is. The calculation formula is shown in equation (2.1):

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

where x and y denote the enhanced and real images, respectively, μ denotes the pixel mean of the image, and σ is the variance of the image. c_1 and c_2 are 0.0001 and 0.0009, respectively.

The standard deviation measures the degree of variation in the pixel values of an image and assesses the contrast and sharpness of the image. Here pixel mean can be used to measure the overall brightness of an image, the higher the pixel mean, the sharper the image. Suppose that there is a data set of $X_i, \quad i \in \{1, 2, \ldots, n\}$ The standard deviation of this data set is:

$$std = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2}. \quad (2)$$

2.5 Pre-processing results

(i) Image denoising

Upon comparing a portion of the original image with the denoised image as illustrated in Fig.6:, it is evident that the denoised apple image exhibits greater clarity than the original image. Additionally, the shape contour is more distinctly separated from the background in the denoised image.

(ii) Image Enhancement Processing

It can be seen in Fig.7that the difference between the apple and the background after image enhancement is obvious, and at the same time, it can better retain the detailed information in the original image to generate a higher quality image, and does not appear more serious distortion.

(iii) Image conversion to RGB format

As can be seen from Fig.8, there is little difference between the RGB-converted image and the original image in terms of sharpness and contrast.

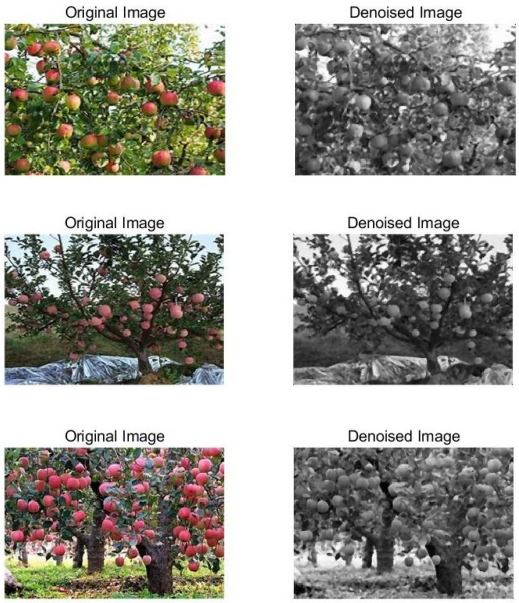

Figure 6: Image denoising

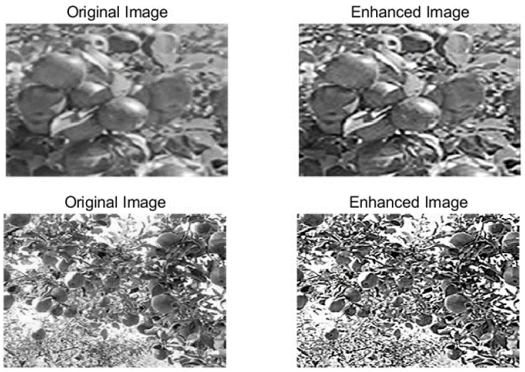

Figure 7: Comparison graph after image enhancement process

Figure 8: Comparison of original image and RGB image

2.6 Image Processing Evaluation

In order to quantitatively analyze the three preprocesses, the SSIM values of each of the three preprocesses were calculated during the experiments to evaluate the enhancement of the im-

ages. The experimental results are shown in Table 1, where the data are selected from the dataset in Subset 1.

Table 1: SSIM metrics for three preprocessing methods

Preprocessing methods	SSIM
Denoising	0.5143
Image enhancement	0.83345
RGB	0.13506

Table 1 lists the results of structural similarity calculations for different treatments, from which it can be seen that the quality of the generated images is improved by using the three preprocessing to enhance the apple images. In terms of SSIM metrics, image enhancement improves 0.319 and 0.698 compared to the two models of direct grayscale processing and image denoising, indicating that the images generated by the image enhancement process are less affected by noise, and the visual effect and brightness of the images are significantly improved.

Table 2: Standard deviation comparison

Preprocessing methods	Average ixels values of the original image	Average pixel value of processed image
Denoising		112.39
Image enhancement	106.34	112.84
RGB		110.34

Table 2 lists the standard deviation results of the different treatments, compared with the original image pixel mean value of 106.34, all have obvious improvement, in which the image denoising and image enhancement in the pixel mean value of the difference of only 0.45, it shows that the image resolution of these two processing methods is higher. After comparison and contrast above, we finally chose the image after image enhancement processing in improving the image brightness, contrast and clarity is better, on the basis of which the apple feature extraction operation is carried out.

2.7 Apple feature extraction based on image processing

(i) Apple circumference

Perimeter is corrected by counting the number of pixels on an object's contour line, in the oblique direction, which produces errors specific to digitized images, by twice their number. When scanning the image from left to right and from bottom to top, a pixel value of 1 is found and its neighboring pixel values (8 neighborhoods) have a different value from it, the apple counter is added to 1, and the entire image is scanned to get the perimeter of the apple.Some of the results are shown in Fig.9.

(ii) Apple color

In this problem, color extraction is done on the basis of image enhancement, so the color features of apples are extracted from the grayscale histograms of the images, and since there

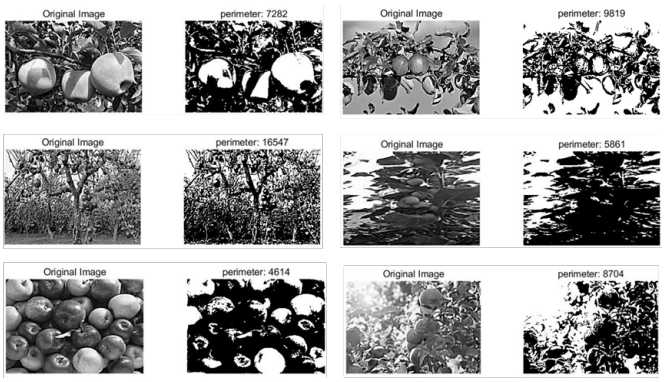

Figure 9: Perimeter features extracted after partial image enhancement

are 200 images in Subset 1, the grayscale values of each image are stored in a matrix. The feature matrix featureMatrix with size (numImages, 256) where numImages is the number of grayscale images in the image folder. This feature matrix holds the normalized grayscale histogram features of each grayscale image. It is shown in Fig.10:

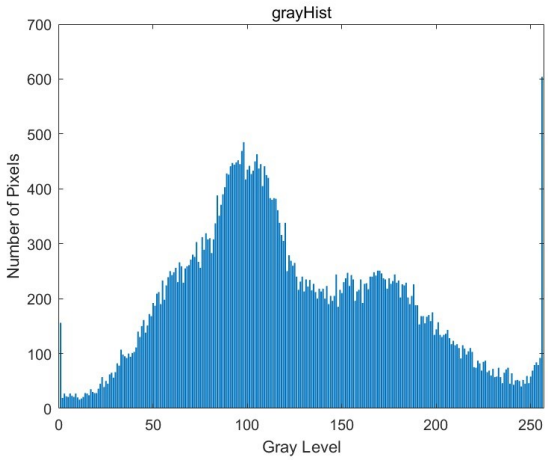

Figure 10: Grayscale histogram

As can be seen from the figure, the gray level of the apple is concentrated in the [50,150] interval and the number of pixels is in the [0,500] interval.

3 Fundamental model

3.1 A counting model for labeling apples based on the labelImg

Before using the YOLOv5 model to detect the position of the apple, we first need to label the position of the apple in the image, the labeling tool used here is labelImg, according to the title of the apple image information given by the use of the edge of the box is labeled, because the image is blurred or foliage obscured by the apple using a manual labeling method,

to improve the accuracy of the number of apples, the position, and to save the information of the labeling.

The data labeling steps are shown in Fig.11: The labeled

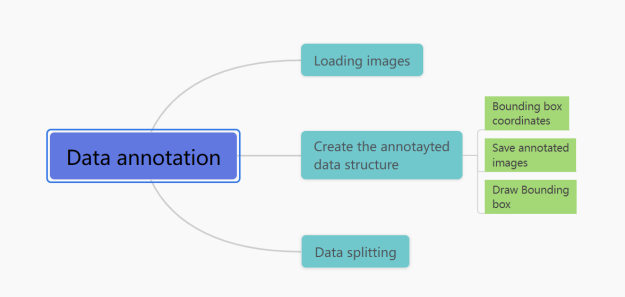

Figure 11: Data annotation content

results are shown in Table 3

Table 3: Number of apples in each image

Image number	Number of apples
1	4
2	7
3	15
4	15
...	...
196	21
197	11
198	52
199	2
200	21
Total number	2921

Note: The parts labeled in red indicate the same number of apples in the image.

3.2 Apple position detection model based on YOLOv5

(i) Model theory

Determination of apple location requires a target detection method, and in order to regressively predict the category and location information of the target object, we use the YOLOv5 target detection model.YOLOv5 uses an end-to-end mechanism to normalize the image and input it into a convolutional neural network. The network structure is mainly composed of four parts: the input, the feature extraction network, the Neck part and the prediction layer. The model structure is shown in Fig.12.

In this, the input side preprocesses the dataset; the feature extraction network performs the slicing operation on the image to achieve downsampling of the image without loss of information; Neck fuses the features of different latitudes; the prediction layer uses CloUzuo as the bounding box loss function, and the non-maximum value suppression algorithm filters the detected target frames. Through the above work, the best result of detecting the apple location is obtained.

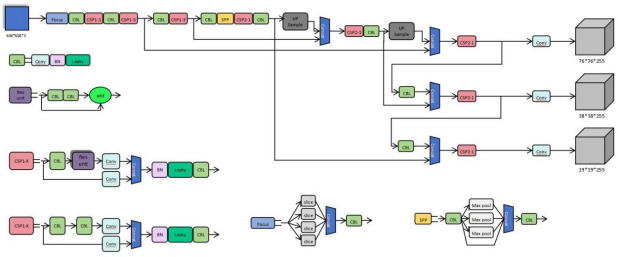

Figure 12: YOLOv5 Object Detection Model Structure

Considering the problems such as foliage occlusion, an ECA attention module is added after the CSP structure of each branch of the YOLOv5s neck network respectively, and the feature information extracted by the feature extraction network is used to perform adaptive learning of features in the spatial dimension to strengthen the fusion ability of features.The feature expression ability of the model in complex scenes is enhanced by adding the attention mechanism module to the neck network, which The interference of irrelevant information is suppressed, so that the model has better detection results in detecting the occluded apple target.

The geometric position of the apple is determined: the center of mass position of the bounding box is measured with the bounding box, and the center of mass position formula is as follows:

$$(x, y) = \left(\frac{\sum x_i}{N}, \frac{\sum y_i}{N} \right), \tag{3}$$

where, x_i, y_i are the coordinates of the pixels in the apple region and N is the total number of pixels in the region.

(ii) Experimental steps

Before using the YOLOv5 model to detect the apple position, we first have to label the position of the apple in the image using the labelImg tool. We use Subset 1 for training and classified Subset 2 for model testing and inspection. The steps are as follows:

Step1: Adding four folders in the data of YOLOv5 directory, Annotations folder is used to store xml files after labeling each image with labelImg; Images folder is used to store the original dataset images that need to be trained in jpg format; ImageSets folder is used to store files used for training, validation, and testing after the dataset has been divided into ImageSets folder is used to store the data set divided into files for training, validation and testing; Labels folder is used to store the labeled files in txt format after converting the labeled files in xml format;

Step2: Preparing the dataset, here we need the geometric position of the apple, so we use all the data in Subset 1 for training, validation and testing;

Step3: Organizing the results of the data obtained from the training and draw a 2D scatter plot of the geometric location;

Step4: Taking the obtained geometric position coordinates and calculate the area of the edge box, which is equivalent to the 2D area of the apple.

(iii) Evaluation system

In this paper, Precision (P) and Recall (R) are used as evaluation metrics to test the model performance. The evaluation metrics are calculated by the formulas respectively:

$$lP = \frac{TP}{TP + FP} \tag{4}$$

$$lR = \frac{TP}{TP + FN} \tag{5}$$

Where TP denotes the number of samples predicted by the model to be positive and are positive samples, FP denotes the number of samples predicted by the model to be positive but are negative samples, and FN denotes the number of samples predicted by the model to be positive but are negative samples.

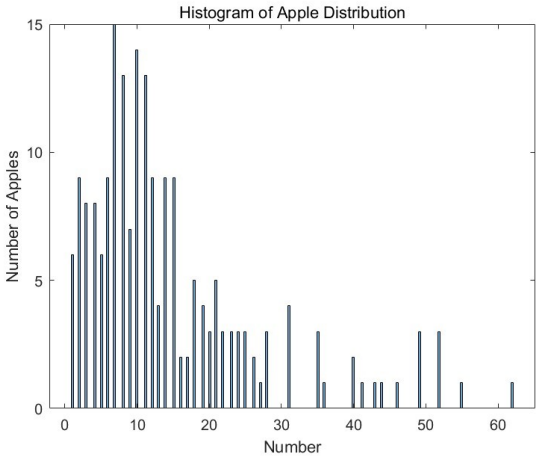

Figure 13: Histogram of Apple Distribution

4 Model prediction results

4.1 Calculate the number of apples

According to the apple image features, we use matlab digital image processing to extract the color and perimeter features of the apple; however, considering that some images have low pixels, which results in the apple features not being obvious in some pictures, the image is first preprocessed before feature extraction, such as adjusting the brightness and contrast to make the apple color and edges in the dark distinctly recognizable, applying filters to remove the noise and performing apply filters to remove noise, color space conversion, and image enhancement to better distinguish apples and background, etc.; after preprocessing, the image is then extracted with its features.

For the calculation of the number of apples, we used labbelmg software to label the data set in Subset 1 according to the given labels, and some apples that were not clear due to the image were labeled and counted manually to get the number of apples. And keeping the number of apples in each image in the labeled dataset, the distribution histogram of all apples was plotted using matlab.

The labeling information is organized into an Excle table to summarize, and then matlab plots the number of apples in each image in Subset 1. The results are shown in Fig.13:

4.2 Estimated Apple Location

The minimum bounding box of the apple is plotted by labbelmg to get the coordinates of the center position of the labeled apple and the length and width size of the labeled bounding box to determine the geometric position of the apple, and the obtained positional data information of the apple is normalized so that the geometric coordinates of the apple can be determined quickly in the next problem of establishing the coordinates to plot the positional information of the apple, and after obtaining the geometric coordinates of the apple, its two-dimensional scatterplot is plotted in matlab to plot its 2D scatter plot. Geometric coordinates of apples Some of the geometric coordinates of apples detected by the YOLOv5 model are shown in Table 4:

Table 4: Geometric coordinates of some apples

Image number	Barycentric coordinate	Bounding box length	Bounding box width
40	(0.890741,0.381081)	0.085185	0.059459
	(0.875926,0.462162)	0.100000	0.059459
	(0.337037,0.681081)	0.074074	0.064865
	(0.246296,0.616216)	0.085185	0.064865
	(0.509259,0.591892)	0.070370	0.059459
	(0.748148,0.275676)	0.066667	0.054054
	(0.753704,0.143243)	0.048148	0.059459
	(0.175926,0.391892)	0.070370	0.048649

The obtained data of bounding box dimensions and center of mass position are normalized and a scatter plot is drawn with the lower left corner as the coordinate origin. The 2D scatter plot of the apple geometric coordinates is shown in Fig.14:

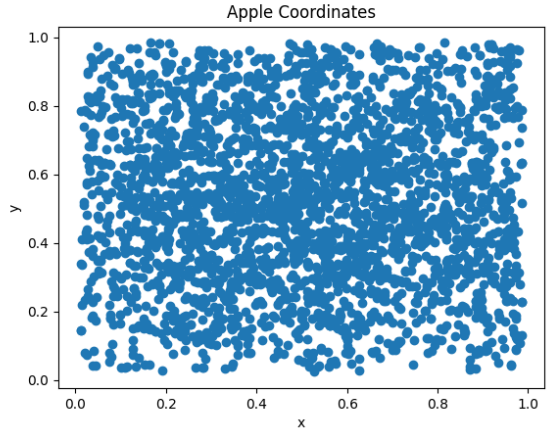

Figure 14: YOLOv5 Object Detection Model Structure

4.3 Estimated quality of apples

The relative area is used to estimate the mass of each apple, in which, for the two-dimensional area of apples, we use the area of the edge frame after the edge detection marking, and the information of the edge frame length and width data obtained to estimate the two-dimensional area of each apple, and the apple mass is estimated according to the average density of apples and the estimation formula. At the same time, taking into account the estimation error impact of the difference between the edge frame of the calculated area being rectangular and the oblate shape of the apple, we provide a certain confidence interval and error range for the estimation results, and produce a histogram of the mass distribution.

According to the mass formula, theoretically, the larger the area, the greater the mass of the apple. Therefore, here we use the ratio of area to mass to solve for the area of the relative image, i.e. the relative area of the apple, to estimate the mass of the apple.

The following figure shows the histogram of apple mass distribution.

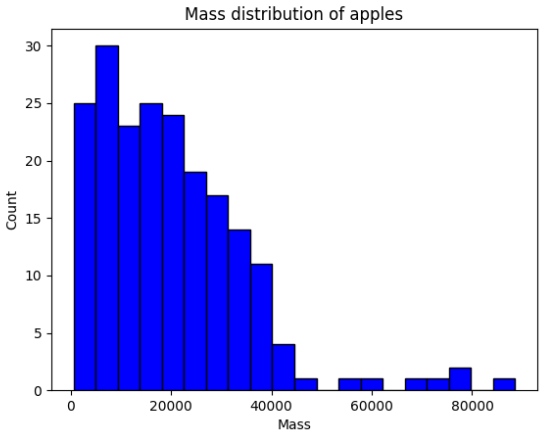

Figure 15: Histogram of apple ripeness distribution

As can be seen from the Fig.15, the apple quality is mainly concentrated in the [0, 2000] interval, and a small portion is distributed in the [60,000, 80,000] interval. The quality and size of the apple images in Subset 1 are basically the same.

4.4 Estimating apple ripeness

Before judging the estimated maturity of apples, we first determine the expression of apple maturity, according to the literature and common sense in daily life, we know that the color, size and texture can show the maturity state of apples, for example, lime green apples are immature and red apples are usually in a ripe state, by using these factors to determine the maturity of apples and coding the different maturity levels, and deepening the prominence of the image features to classify the ripeness of apples, again, we use YOLOv5 detection model for classification.

The maturity of apples is related to many factors and belongs to the multi-classification problem, which is usually categorized into different maturity levels, divided into four levels, and coded in the data annotation. The coding is defined as follows:

Table 5: Apple Data

Coding	Apple color	Tag number	Grade of maturity
1	All-red	15	mature
2	All green	17	Immature
3	Half red and half green	16	Semi-mature
4	(flower)bud	19	Extremely immature

Based on this criterion, the classification results are shown in Fig.16:

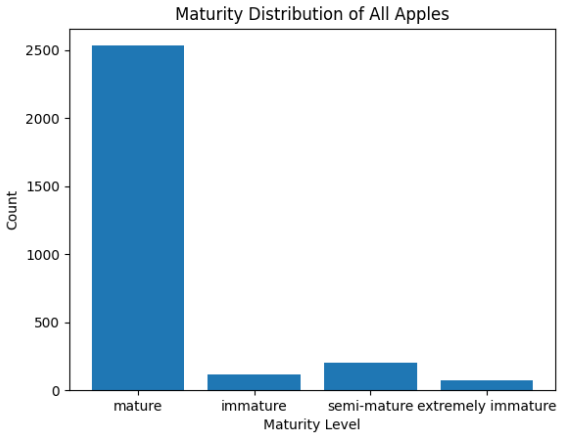

Figure 16: Histogram of apple ripeness distribution

As can be seen from the graph, the largest percentage of fully red apples, over 2,500, indicates a majority of fully ripe apples; followed by semi-ripe apples with the fewest flower bones.

4.5 Apple Recognition Based on YOLOv5 Models

We divide the classified labeled Subset 2 data as the training set and use Subset 3 as the test set to train the YOLOv5 detection model to achieve apple detection in Subset 3 data. The obtained results are shown below in Fig.17:

From the figure, it can be seen that [0,4140] has the highest number of ID number apples, close to 1400, and [8280-10350] has the lowest number of ID numbers, only about 800, with precision and recall rates of 99.08% and 99.3%.

5 Conclusions

(i) In feature extraction, the image quality is improved by preprocessing steps such as image denoising, enhancement and color space conversion, which makes the sub-

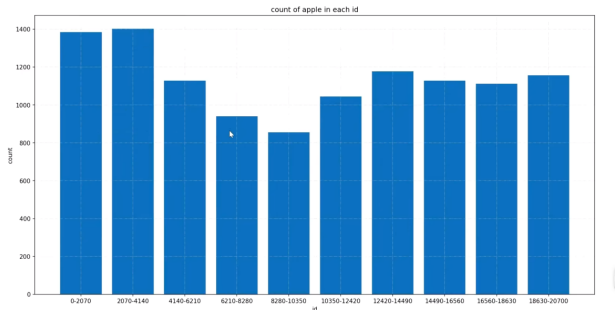

Figure 17: Histogram of apple ripeness distribution

sequent feature extraction and target detection more accurate and reliable.

(ii) Based on the detection model of YOLOv5, enhanced with an attention mechanism module, successfully extracts and recognizes features from robot apple picking images in complex scenes. This enhancement improves the model's feature expression ability, extraction capability of apple image features, and detection speed, enabling accurate detection of apple locations, numbers, maturity, and size.

(iii) The experimental results show that our proposed method has good performance and accuracy in apple picking tasks, which provides strong support for the development of robotic automated picking systems. Future research directions can further optimize the algorithm and improve the accuracy of detection and recognition to meet the needs of more efficient in real production.

Acknowledgment

The authors are grateful to the editor and the referees for their valuable comments and suggestions. This work was supported by College Student Innovation and Entrepreneurship Training Program Project (S202410608112).

Data availability statement

The data that support the findings of this study are available from Asia-Pacific Student Mathematical Modeling Contest Organizing Committee, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors(corresponding author:229744332@qq.com) upon reasonable request and with permission of Asia-Pacific Student Mathematical Modeling Contest Organizing Committee.

Ethical Approval

Not applicable.

Competing interests

There is no conflict of interests.

Authors' contributions

The authors contributed equally to this paper.

References

[1] SONG Zhe, WANG Hong, LILI Hui, et al,"Main problems, development trend and solutions of apple industry in China," Jiangsu Agricultural Science,vol.44,no.9,pp.4-8,2016.

[2] WANG Dandan, SONG Huaibo, HE Dongjian,"Research progress of apple picking robot vision system," Journal of Agricultural Engineering,vol.33,no.10,pp.59-69,2017.

[3] KAPACHK,BARNEAE,MAIRONR,etal,"Computervision for fruit harvesting robots-state of the art and challengesahead," International journal of computational vision androbotics,vol.3,no.2,pp.4-34,2018.

[4] Cao Chunqing, Zhang Wuping, Li Fuzhong, et al,"Research on fusion algorithm for multi-target apple recognition and localization in natural scenes," Hubei Agricultural Science,vol.61,no.7,pp.145-151,2022.

[5] ZHAO De'an, WU Rendi, LIU Xiaoyang, et al,"Robotic apple picking localization in complex context based on YOLO deep convolutional neural network," Journal of Agricultural Engineering,vol.35,no.3,pp.164-173,2019.

[6] Cao ZP,"Research on actuator and target detection of apple picking robot," Kunming University of Science and Technology,2023.

[7] Wang Yong, Tao Zhaosheng, Shi Xinyu, et al,"Target detection method of different maturity apples based on improved YOLOv5s,"Journal of Nanjing Agricultural University,2023.

[8] ZHANG Shifu,"Research on apple target recognition and localization algorithm based on deep learning," Zhejiang University of Technology,2020.

[9] Song Yang,"Research on image enhancement and apple detection method based on deep learning," South-Central University for Nationalities,2022.

[10] ZHANG Tao; LI Zhisheng,"Apple object detection based on BG and RTHTR image processing," Electronic Design Engineering,vol.31,no.10,pp.135-140,2023.

[11] HU Shilin, CHEN Wei, ZHANG Jingfeng, et al,"Target detection method of apple picking robot based on improved YOLO v5," Agricultural Mechanization Research,2024.