

Research on apple detection and maturity assessment based on computer vision technology

Junwen Zheng[#], Yuhui Zhang^{#, *}, Ying Wang[#]

College of Information and Intelligence Engineering, Zhejiang Wanli University, Ningbo, China, 315100

*Corresponding author: tensor0404@gmail.com

[#]These authors contributed equally.

Abstract. The study is based on computer vision technology for apple detection and maturity assessment. Firstly, the YOLOv7 algorithm is used for target detection of apples, and the number of apples in the image is counted to generate a distribution histogram of the number of apples. Then, the position of each apple is detected by the YOLOv7 algorithm and a 2D scatter plot of the geometric coordinates of the apples is drawn. Next, apple foreground was extracted by interactive ROI tagging and GrabCut algorithms, and a mathematical model was developed to assess the maturity of apples based on the histogram analysis of apple colours in HSV colour space model. In addition, the Faster R-CNN model is used to detect apples and estimate the quality of apples based on the 2D area of apples to generate a distribution histogram of apple quality. Finally, a convolutional neural network is used to build a fruit recognition model and draw a distribution histogram of apple image ID numbers. This study provides computer vision technical support for apple picking robots to improve apple production efficiency.

Keywords: Computer vision; Apple detection; Ripeness assessment; YOLOv7; Faster R-CNN.

1. Introduction

China's annual apple production has reached 35 million tons, making it the world's largest apple exporter. Meanwhile, in the context of the Belt and Road Initiative (BRI), Chinese apples have also gained strong demand from the markets of the countries along the route. Therefore, it is urgent to improve the efficiency of the apple production process, and the Chinese government has been working on the development of apple picking robots to address the huge labor shortage during the picking season in labor-dependent apple-producing regions and to improve the economic efficiency at the same time. Significant progress has been made in this research, but there are still some general applicability challenges with these technologies due to the diversity and complexity of orchard environments. In order to improve the mechanization of apple picking and truly achieve the purpose of helping farmers, this paper will establish a mathematical model to solve the key problems of apple quantity, location, ripeness, quality, etc., improve the model generalizability, optimize the image recognition capability of the fruit picking robot, and promote the development of the apple industry.

This paper intends to address the following questions:

- (1) Extract key features from the image dataset, construct a mathematical model to calculate the number of apples in each image, and produce a histogram to show the distribution of the number of all apples.
- (2) Identify and locate the position of each harvestable apple by using the lower left corner of each image as the origin of the coordinate system and produce a 2D scatter plot showing the geometric coordinates of the apples.
- (3) Determine the feature conditions, construct a mathematical model from the image dataset of ripe apples, assess the degree of ripeness of the apples in each image, and produce a histogram to show the distribution of ripeness.

(4) Using the dataset of images of ripe apples, an algorithm measures the two-dimensional area of the apples in each image, estimates the mass of the apples using the lower left corner of the image as the starting point of the coordinate system, and creates a histogram to show the distribution of mass.

(5) Extract key features from the image dataset, train an apple recognition model to accurately recognize the apples, and produce a histogram to show the distribution of ID numbers.

2. Research methods

2.1. Apple target detection and counting

It is envisioned to realize apple target tracking as well as counting with the help of relatively robust and automated YoLov7 target detection algorithm. Firstly, the data samples were analyzed and found to have small sample capacity and low clarity. At the same time, the environmental background of the pictures was observed, and there were more interference items, such as the light and shadow interference generated by the difference of various light sources in nature, among which the samples of green apple varieties mixed with green leaves were the most special, and in addition to the common interference items, the differentiation of their color features was also relatively inconspicuous. Since light sources and backgrounds affect recognition, it is first envisioned to preprocess pictures to enhance boundary clarity. In general there may be size differences in the picture, but after confirming that the size of the picture is uniform, then there is no adjustment in this paper. In response to the conventional way of finding boundaries using color feature differences, it is envisaged to use color space models such as hav, yuv, hsi, etc., to separate the color information from the luminance information, but after observing and comparing the recognition effect in YoLo-v7 is not ideal. Therefore it is envisioned to process the image using image denoising basic method of filtering as well as expansion corrosion. Test set test is applied and if the number of recognized images is accurate after sampling and comparison, the model is reasonable. The preprocessed images are then fed to obtain a table of the number of apples in each image. According to the topic plotting requirements, it is envisioned that the table will be plotted as a histogram with the number of apples as the x-axis and the frequency of occurrence of the quantity as the y-axis, taking into account the need for orchard yield prediction.

2.2. Apple location localisation

The same image preprocessing is envisioned. Since the YoLov7 algorithm not only detects the target, but also performs target localization providing information about the target's location, including the coordinates and size of the bounding box, it is envisioned that it is modified on the established pre-existing model to estimate the location of the apples in the image. Due to the complexity of the image background, it is envisioned to use a non-great value suppression algorithm commonly used in image recognition to remove redundant information in order to retain the most relevant results. A test set test is applied and the model is reasonable if the number of recognized images are sampled and compared with high accuracy. Then input the preprocessed image, envision to get the location information two points coordinates of the upper left and lower right corners, and calculate the center value using the formula. Finally output the table to draw a 2D scatter plot of the apple set coordinates.

2.3. Apple maturity assessment

In accordance with the requirements, the characteristic distinction of apples at different periods was consulted, in which the most obvious characteristic of the skin visible in the image is the color. It is envisaged to differentiate by color features, i.e., to analyze the color distribution of each apple and to set a criterion for the degree of ripeness. It is envisioned that the data will be divided into several different discrete categories rather than classified according to a continuous degree, based on the fact that this paper is about fruit picking robots, which are ultimately categorized into four categories: pickable ripe apples, non-pickable half-ripe apples, unripe apples, and blossoms. It has

obtained some of the positional coordinate information of the apple, it is envisioned to input the excel coordinate table to frame out the apple region with the help of interactive roi, and to perform foreground segmentation with matlab's grabcut algorithm, because this algorithm only needs to specify a rough border that can frame the target to accomplish a good segmentation. And then envisage the use of hsv color space model, so that the segmented apple image separated from the color histogram made of color histogram analysis to determine the maturity, and finally arrived at apple maturity distribution histogram.

2.4. Apple Quality Detection

The need for image recognition capability is taken a step further and it is envisioned that a Faster R-CNN deep learning model is used to make the model more generalizable. After getting the trained model, since the model can output the bounding box of the object and the corresponding confidence score, it is envisioned to set a threshold, and then go for volume estimation for objects above the threshold, and thus estimate the mass. Finally a histogram is plotted to show the apple mass distribution.

2.5. Apple Recognition

A more sophisticated one with high generalizability is needed, and it is envisioned to use the sensorFlow and Keras libraries to build a deep learning model, which is trained by a convolutional neural network (CNN) model on a dataset that contains a variety of fruit categories. Finally, a histogram of the distribution of all apple image ID numbers is plotted.

3. Results and analyses

3.1. Apple target detection and counting

3.1.1 Image pre-processing

The image preprocessing process in this paper is represented by a flowchart as follows(Figure 1).

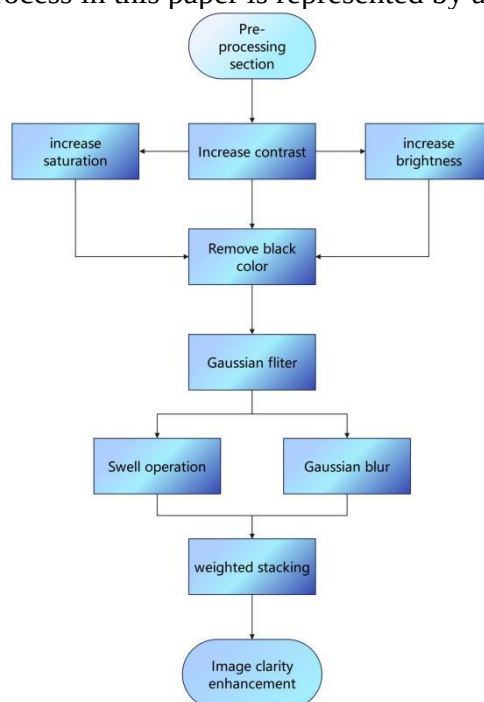


Figure. 1 Image Processing Flowchart

Images taken in natural environments are often affected by various forms of noise that degrade the quality of the image and also reduce the contrast between the apples and the background. This

phenomenon can blur the line between the apples and the background, making it difficult to accurately identify and distinguish the apples.

In this paper, we will first gain the contrast, saturation, and brightness of the image using the codes $\alpha = 1.2$, $\beta = 10$, and $\gamma = 1.2$.

In the second step, use the `cv2.inRange()` function to create a mask that sets the positions with pixel values greater than 0 (the positions where the mask is white) to [60, 60, 60], and then corrects the pixel values in the image that are less than [0, 0, 0] or greater than [60, 60, 60] to [60, 60, 60], so that the pixels are converted to white (255), which means that the black dots are removed, making the apple image more in line with the differentiation requirements.

In the third step, a variety of filtering methods were tried, and finally the Gaussian filtering process was optimized to remove the noise and suppress the noise in the image for each picture (Figure 2).



Figure. 2 Comparison of original and median, Gaussian, and bilateral filtering

In the fourth step, the `cv2.dilate` function in the OpenCV library is used to perform an expansion operation on the image to increase the resolution, and then the Gaussian Blur function is used to perform Gaussian blurring on the image to reduce the noise in the image and reduce the details. Then the `cv2.addWeighted` function is used to perform a weighted superposition of the blurred image with the original image and adjust the weighting parameters to enhance the clarity of the image.



Figure 3 Take 47.pgn as an example, the left is the original image and the right is the processed image

The above processing helps to subsequently better identify and analyze the apple edges and features in the image and improve the model robustness (Figure 3).

3.1.2 Modeling apple counts

Apple target tracking using YoLov7 target detection algorithm to output data. The following is the structure of the YoLov7 network (Figure 4) [1-4]:

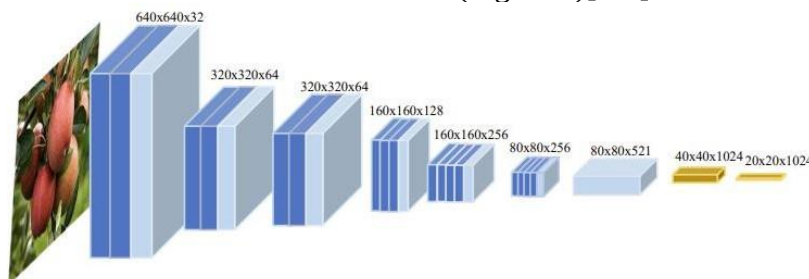


Figure 4 YoLov7 network structure

The image was entered for detection and the results are shown in the legend (Figure 5).



Figure 5 Detection results

3.1.2 Problem solving

The data will be output excel table, the use of excel software plotting function for histogram plotting, the number of apples in a single figure (data is divided into five intervals) as the x-axis, the frequency of occurrence of the y-axis of the distribution of the histogram, and increase the total percentage display to make the graph clearer(Figure 6).

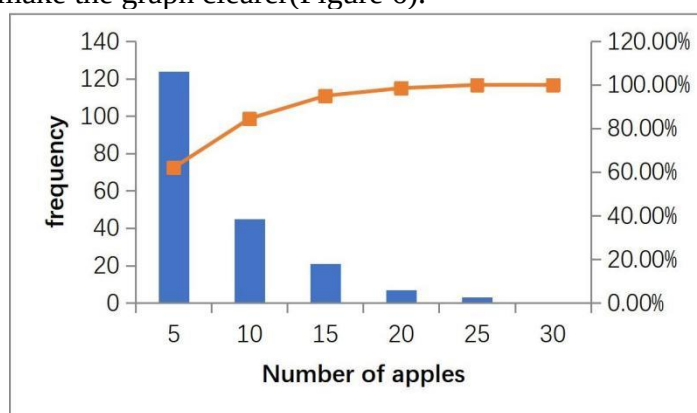


Figure 6. Histogram of apple number distribution

3.2. Apple location localisation

3.2.1 Apple Location Estimation Models

Target detection is first performed in Python based on the yolo-v7 model, which is described in detail above and will not be repeated. After detecting the apple image, redundant detection frames with high overlap are removed using the Non-Maximum Suppression (NMS) algorithm[5,6].

3.2.2 NMS algorithm Redundancy removal steps

Set the confidence threshold and remove those detection boxes whose confidence level is below the threshold.

Select the frame with the highest confidence level from the remaining test frames as the reference frame.

Calculate the intersection and concurrency ratio (IoU) of all other boxes to this highest confidence box remove all frames whose IoU with the reference frame exceeds a specific threshold; these highly overlapping frames are likely to be multiple detections of the same target.

Get the coordinates of the lower left and upper right corners of the border and calculate the center point coordinates.

The center point coordinate formula code is:

$$mid_x = (x1 + x2) / 2.0 \quad mid_y = (y1 + y2) / 2.0 \quad (1)$$

3.2.3 Problem solving

A 2D scatterplot of all the geometric coordinates of the apples was plotted using Python's Matplotlib data visualization library(Figure 7).

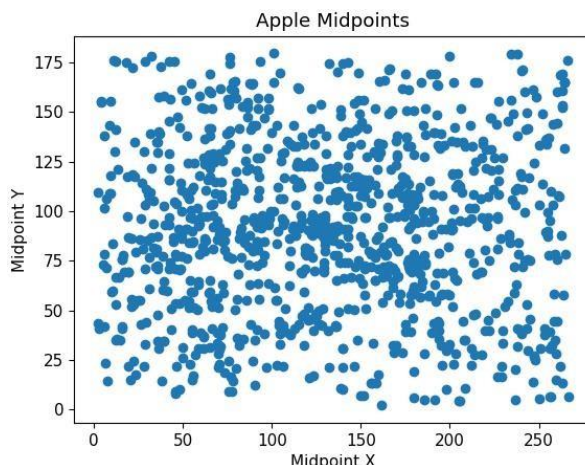


Figure 7. 2D scatterplot of all apple geometric coordinates

3.3. Apple maturity assessment

3.3.1 Image Preprocessing

Based on the set of image coordinate information, spatial difference calculation is performed in Excel to obtain the complete quadrilateral position coordinates, i.e., $(x1, y1)$, $(x2, y2)$, $(x3, y3)$, $(x4, y4)$.

Table 1. Example of coordinate information

Picture	Number	x1	y1	x2	y2	x3	y3	x4	y4
1	1	138	73	138	103	166	103	166	73
1	2	119	1	119	29	155	29	155	1
1	3	106	38	106	91	160	91	160	38
2	1	36	117	36	124	42	124	42	117
2	2	65	148	65	155	73	155	73	148

The data is fed into Matlab, the block diagram is obtained using interactive Rio marking the coordinate positions, GrabCut algorithm image segmentation, and extraction of foreground object apple (Table 1).

3.3.2 Interactive Rio

In MATLAB, interactive RIO (Region of Interest) marking refers to the use of MATLAB's tools and functions to interactively mark and analyze specific regions (regions of interest) in an image during image processing and analysis. Once the ROI is selected, pixel data from that region can be extracted for further analysis or processing. The RIO method mentioned in this paper i.e. marking the range of pixels where apples are located by inputting coordinate values for image background segmentation to reduce the interference terms.

3.3.3 GrabCut algorithm

The algorithm uses the color information and boundary information in the image to extract the foreground object from the background, and a relatively good segmentation result can be obtained with a small amount of user interaction. As shown in Figure 8.

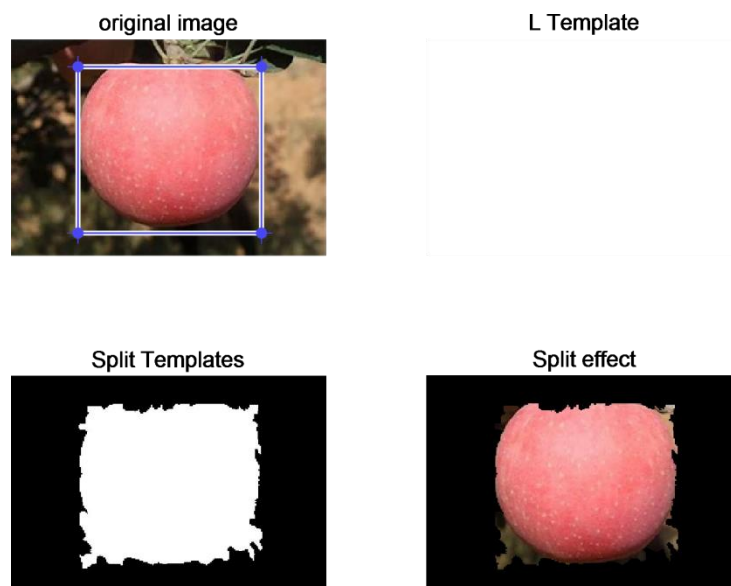


Figure 8 Example of processing 110.jpg

3.3.4 Modeling

Ripeness was classified into four categories: pickable ripe apples, non-pickable semi-ripe apples, unripe apples, and blossoms, and the color histogram was divided into four intervals to determine ripeness.

After extracting the segmentation effect image, it was transformed into HSV color space model, separated the three-channel components, and selected the distribution characteristics in terms of hue (H) to make a color histogram for analysis.

Values of H less than 1 are classified as flowers, when the value of h is greater than or equal to 1 and less than 5 are classified as unripe apples, when the value of h is greater than or equal to 5 less than 15 are classified as unpickable semi-ripe apples, and the rest of the values of h greater than 15 are classified as pickable ripe apples.

3.3.5 Problem solving

The maturity discrimination results were entered into Python to integrate the data to produce a histogram of the maturity distribution of all apples(Figure 9).

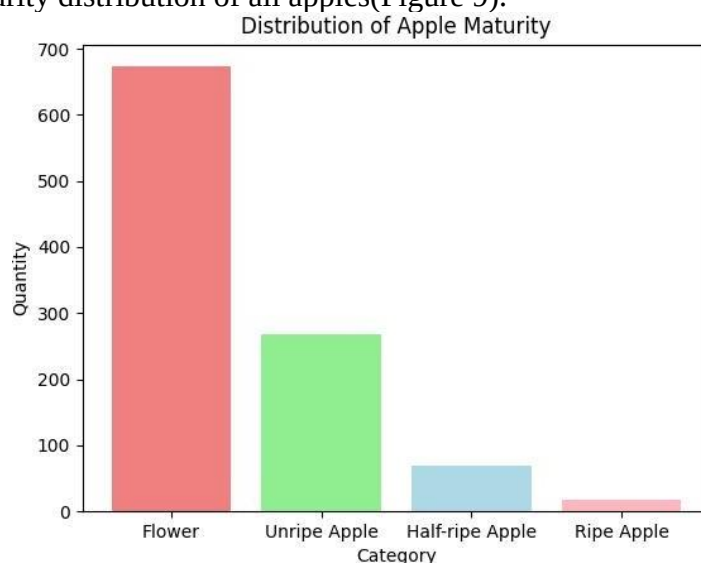


Figure. 9 Histogram of maturity distribution of all apples

3.4. Apple Quality Detection

3.4.1 Image Preprocessing

Consistent with section 3.1 Image Preprocessing.

3.4.2 Modeling

Pre-training in Python using the deep learning model Faster R-CNN model to obtain an optimized apple detection model that is more generalizable and robust.

3.4.3 Faster R-CNN models

Faster R-CNN (Region-based Convolutional Neural Networks) is an efficient target detection model proposed in 2015 by Shaoqing Ren, Kaiming He, Ross Girshick and Jian Sun. The model improves on the previous R-CNN and Fast R-CNN, and accelerates the target detection process mainly by introducing Region Proposal Network (RPN) to achieve higher accuracy[7,8].

3.4.4 Problem solving

Object detection is performed using the trained model, the bounding box and the corresponding confidence score for each detected object are output, and for objects with a confidence level higher than a set threshold, the 2D area is computed, and the volume and mass are estimated using conventional formulas. A histogram was plotted to show the distribution of the quality of all detected apples(Figure 10).

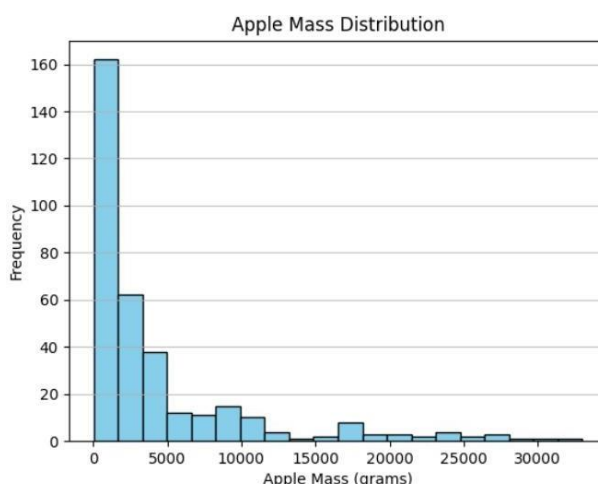


Figure. 10 Histogram of the mass distribution of all detected apples

3.5. Apple Recognition

3.5.1 Image Preprocessing

In this study, the fruit harvest image dataset was first preprocessed and enhanced using ImageDataGenerator for TensorFlow. Data enhancement techniques such as normalization, miscutting, scaling, and horizontal flipping are employed with the aim of improving the robustness and generalization of the model.

3.5.2 Modeling

We constructed a convolutional neural network (CNN) based deep learning model for recognizing fruit categories. The model consists of convolutional, pooling and fully connected layers, as well as an output layer (softmax activation layer for 5 categories). The model was trained for images of 150x150 pixel size for a total of 10 epochs, using the adam optimizer and categorical cross-entropy as loss functions[9,10].

3.5.3 Problem solving

We use the trained model for object detection and output the bounding box and the corresponding confidence score for each detected object. For objects with a confidence level higher than a set

threshold, we calculate their 2D area and use a generalized formula to estimate their volume and mass. In addition, we generate a histogram to show the mass distribution of all detected apples to better understand the quality characteristics of the detected apples. This process is typically used for computer vision tasks and can automatically detect and estimate object properties.

4. Conclusions

In this study, the problem of apple detection and maturity assessment was successfully solved by computer vision technology. The YOLOv7 algorithm achieved the detection and localisation of apple targets, and generated the distribution histogram of the number of apples. The GrabCut algorithm was used to extract the apple foreground, and combined with the HSV colour space model of apple colour histogram analysis, the assessment model of apple maturity was established. In addition, the Faster R-CNN model was implemented to detect and evaluate the apple quality. This study provides effective computer vision technical support for apple picking robots and lays the foundation for improving apple production efficiency. Future research can further optimise the model and improve its generalisation ability to adapt to more complex orchard environments.

References

- [1] Ye G ,Li S ,Zhou M , et al. Pavement crack instance segmentation using YOLOv7-WMF with connected feature fusion[J].Automation in Construction,2024,160105331.
- [2] Chen T ,Li H ,Chen J , et al. Detection network for multi-size and multi-target tea bud leaves in the field of view via improved YOLOv7[J].Computers and Electronics in Agriculture,2024,218108700.
- [3] Li Z ,Zhu Y ,Sui S , et al. Real-time detection and counting of wheat ears based on improved YOLOv7[J].Computers and Electronics in Agriculture,2024,218108670-..
- [4] Zhou H ,Wu S ,Xu Z , et al. Automatic detection of standing dead trees based on improved YOLOv7 from airborne remote sensing imagery[J].Frontiers in Plant Science,2024,15.
- [5] Danyu W ,Zhen L ,Xingyu G , et al. Feature extraction and segmentation of pavement distress using an improved hybrid task cascade network[J].International Journal of Pavement Engineering,2023,24(1).
- [6] Jiang T ,Sun Y ,Huang H , et al. Binocular Vision-Based Non-Singular Fast Terminal Control for the UVMS Small Target Grasp[J].Journal of Marine Science and Engineering,2023,11(10).
- [7] Li J ,Wei X .Research on efficient detection network method for remote sensing images based on self attention mechanism[J].Image and Vision Computing,2024,142104884.
- [8] Jixuan H ,Xueqin H ,Zhaoli D , et al. IoU-aware feature fusion R-CNN for dense object detection[J].Machine Vision and Applications,2023,35(1).
- [9] Yan X ,Liu X ,Li J , et al. Recoding double-phase holograms with the full convolutional neural network[J].Optics and Laser Technology,2024,174110667.
- [10] Kazama H E ,Tedesco D ,Carreira S D V , et al. Monitoring coffee fruit maturity using an enhanced convolutional neural network under different image acquisition settings[J].Scientia Horticulturae,2024,328112957.