Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

A smart fruit size measuring method and system in natural environment

Bingkai Wang, Mengqi Li, Yuqi Wang, Yuhan Li, Zilan Xiong *

State Key Laboratory of Advanced Electromagnetic Technology, Huazhong University of Science and Technology, WuHan, Hubei, 430074, PR China

ARTICLE INFO

Keywords:
YOLOv5
Size measurement
Object detection
Occlusion recovery
Cycle GAN

ABSTRACT

The measurement of fruit size (length, width, and area) is one of the key components in the assessment of fruit ripeness for consumption and fruit quality. Due to issues with fruit shape and ambient occlusion, noncontact measurement of fruits in their natural growth state in natural environment is more difficult. This study proposes a fast and low-cost noncontact measurement solution combining deep learning, image analysis and robotic platform. First, based on the YOLOv5 object detection algorithm and converged dataset, a 95.6% detection accuracy and 0.05 s detection speed was achieved. Second, a Cycle GAN model for fruit occlusion recovery was established, and the occlusion of the fruit under different natural conditions was adaptively repaired, with a 5.48% average relative error under different occlusion conditions. Third, through image morphological operations, the size measurements were achieved, with an average relative error of 8.43% for individual fruits and the 10.12% overall error under the complex natural environment. This method and platform provide a systematic, fast (0.2s) and low-cost solution for the noncontact measurement of fruit size.

1. Introduction

The measurement of fruit size in their natural state plays a crucial role in the assessment of food quality. By quantifying the dimensions of fruits in their natural environment, researchers and farmers can obtain essential information regarding the maturity stage and physiological changes occurring within the fruit. Fruit size measurement parameters, including length, width, area, volume, and weight(Miranda et al., 2023). Among the parameters, area, length, and width are the fundamental fruit dimensional parameters that are most likely to be directly measured from the image. But the growth status of fruits in natural environments is complex, and monitoring the entire growth cycle of fruits requires a measurement approach that is fast and low-cost.

To measure fruit size, the detection of fruits must be realized in the natural environment first. Common detection methods include digital image (Donmez et al., 2021; Gan et al., 2020; Guo et al., 2020) and machine learning(Fan et al., 2020; Sari and Gofuku, 2023; Xu et al., 2024). However, the accuracy is limited in natural environments and can easily be affected by factors, such as occlusion. However, occlusion is a major problem when measuring fruit size under natural conditions. Improved Hough Transform (Chen et al., 2021) and deep learning-based occlusion recovery methods (Ge et al., 2019; Kim et al., 2023; Magistri et al., 2022) were proposed to solve the occlusion. However, these methods cannot be used for noncircular fruits, or with slightly low precision, or increased costs. In recent decades, digital image processing techniques have achieved significant application in fruit size measurement. Common image processing algorithms have been applied to individual fruits(Al-kaf et al., 2020; Lu et al., 2022; Phate et al., 2021). Multi-image or multi-camera solutions and 3D RGB-D sensors, have made progress in accuracy and ability for capturing the three-dimensional characteristics of a fruit(Jadhav et al., 2019; Liong et al., 2023); however, the cost is high and it is not suitable for natural conditions.

In the abovementioned research on fruit size measurements, their limitations are summarized as follows: (1) The detection of fruits must be carried out in a complex natural environment, with the detection speed and accuracy considered simultaneously. (2) The existing occlusion recovery schemes are either constrained by fruit shape, or have slightly low accuracy or high cost. (3) Existing image measurement schemes are primarily intended for isolated fruits in a fixed position, which are not suitable for natural conditions. To address the challenges in fruit size measurement mentioned above, deep learning methods have been attempted to achieve object detection and occlusion recovery of fruits in natural environments. Combined with image processing techniques and robotic platform, size measurement is achieved. The main contributions of this study are as follows:

* Corresponding author.
E-mail address: zilanxiong@hust.edu.cn (Z. Xiong).

https://doi.org/10.1016/j.jfoodeng.2024.112020
Received 17 November 2023; Received in revised form 22 February 2024; Accepted 25 February 2024
Available online 26 February 2024
0260-8774/© 2024 Elsevier Ltd. All rights reserved.

(1) A fast and light-weight fruit detection model was built using YOLOv5, which combines a self-built dataset and an Open Images dataset to detect fruits accurately.
(2) An occlusion recovery method was developed based on Cycle GAN to realize the adaptive recovery of the natural occlusion of fruits, and it was not limited by the shape of the fruit.
(3) Morphological operations were performed on the fruit image to obtain the pixel size and estimate the actual size of the fruit using the distance value obtained from a ranging device mounted on a robot platform.
(4) A robot platform integrated with different parts was built, and field experiments were conducted to test the performance of the proposed method in complex natural environments.

The rest chapters of the paper are arranged as follows: Section 2 describes the experimental setup, methodology, and evaluation metrics. Section 3 presents the main test results and relevant analysis discussions. Finally, Section 4 provides a brief summary and outlines future direction for improvement.

2. Materials and methods

2.1. The test platform and overall flow-chart of the proposed method

To verify and test the proposed method, a single camera, a ranging device, and wireless transmission were used to achieve automatic image acquisition. As shown in Fig. 1a, for fruit plants of different heights, a camera and a ranging device were mounted on an unmanned aerial vehicle (UAV) or robot to capture pictures of fruit plants at a close distance. The collected images and distance information were sent to a computer in real time via a wireless communication equipment. The brief process of running this operation in the field involves setting the UAV or robot to a specific action program. It will stop at preset position and capture images and distance information, which are then transmitted to the computer. To achieve this, a robotic platform was built to assist with the acquisition of images and distance information, as shown in Fig. 1b. A data collection box was placed next to the plant to gather information of the plant's growth environment and assist in positioning the robot accurately in front of the plant. The data collection box combines the robot's inertial navigation and GPS navigation can achieve a positioning error ~2.5 cm, which help robot to obtain clear fruit image and distance (Zilan Xiong et al., 2021). The image acquisition was completed using a camera (WXSJ-1080P-AHD) controlled by a Raspberry Pi 3B, while the distance information was obtained using a ranging device (HC-SR04) controlled by STM32, and these modules are integrated on the robot. The collected image and distance information were sent to a computer (DELL, Intel i5 9th Gen, NVIDIA 1650) in real time through a wireless network. The stay time of the data collection and transmission process is within 5 s. The running programs were based on Python (Anaconda, CUDA10.0, Cudnn7.8), and the related training and testing were based on this configuration. This study focused on the proposed method for measuring fruit size in natural environments. The image and distance acquisition processes can be adaptively selected according to the actual situation and not the exact example given in

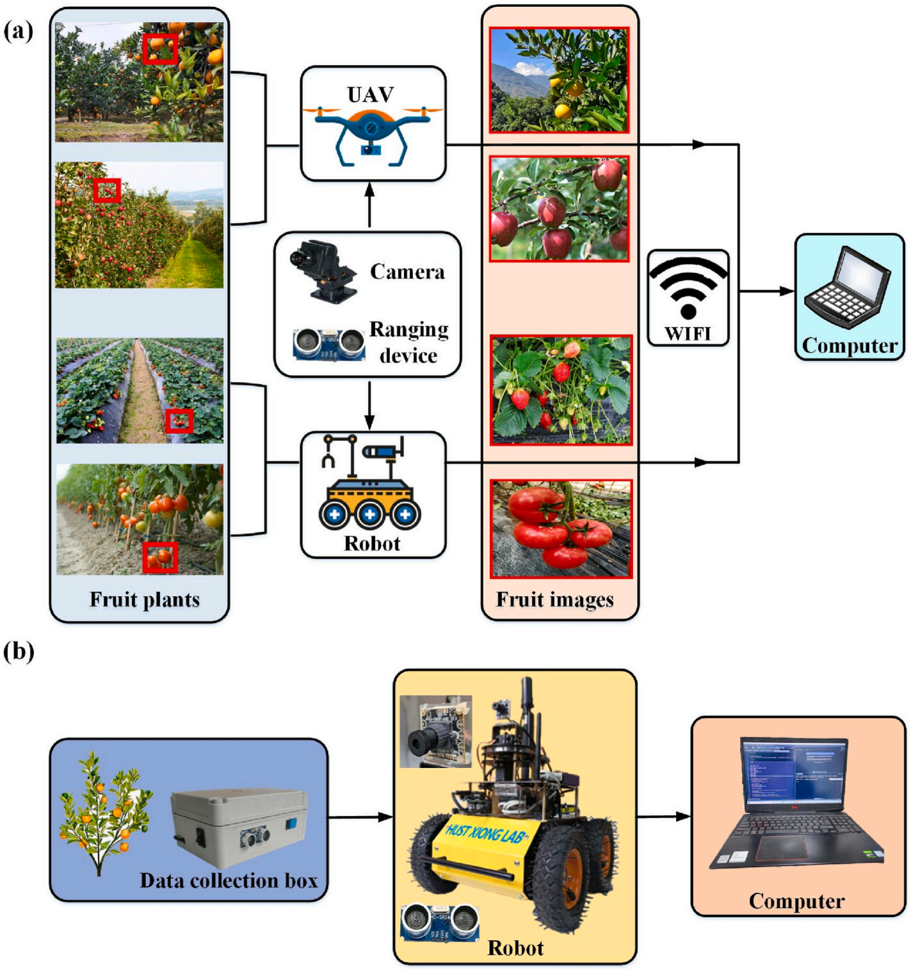

Fig. 1. The main experiment process of proposed method. (a) Automatic image acquisition process achieved by a single camera, ranging device and wireless transmission; (b) The robot platform for image acquisition used in this study.

Fig. 1b.

After the collected images were sent to a computer, the processing flow shown in Fig. 2 was performed. The object detection model is first used to achieve fast and low-cost fruit detection. Subsequently, the color feature of the fruit is used to adaptively repair the occluded fruit image. Finally, the image morphology operation extracts the fruit pixel size information, combines it with the distance information to convert the pixels to the actual size, and obtains the size measurement.

2.2. YOLOv5

Currently, common object detection algorithms are divided into two categories: one-stage and two-stage object detectors. The most representative algorithm for a two-stage object detector is the R–CNN series (Hmidani and Ismaili Alaoui, 2022). The representative algorithms for region-free one-stage object detectors are the SSD(Zhang et al., 2021) and YOLO series(Bochkovskiy et al., 2020; Redmon and Ali, 2018). Among these algorithms, the YOLO series has been optimized and developed in recent years and is generally superior to other algorithms in terms of both detection accuracy and speed. YOLOv5(https://github.com/ultralytics/yolov5) is an object detection algorithm capable of detecting and classifying objects in real-time. YOLOv5 has advanced from the YOLO family of object detection models, due to its high accuracy, speed, and ease of implementation. The YOLOv5 algorithm has been tested on the COCO dataset, where it has achieved high mAP50 and FPS scores exceeding those of other state-of-the-art object detection algorithms.

The YOLOv5 algorithm uses a new network architecture and multi-scale prediction techniques to improve object detection capabilities. The algorithm is divided into four parts: Input, Backbone, Neck, and Prediction. Resizing the input image to 640*640*3 and applying a series of data augmentations help improve model generalization ability. The Backbone employs deep convolutional neural networks to extract high-level features from images, and the Neck enhances feature representation through the incorporation of contextual information and spatial details that are passed to Prediction. After non-maximum suppression and other post-processing techniques, the object detection results are obtained. YOLOv5 employs a fused loss function:

$$Loss_{total} = l_{box} + l_{cls} + l_{obj} \tag{1}$$

Where the l_{box}, l_{cls} and l_{obj} are the box loss function, classification loss function, and confidence loss function, respectively (Feng et al., 2023). The overall loss is the sum of these three losses, and by adjusting the weight value in each function, the attention to the loss of the three can be adjusted to improve the generalization ability of the model.

2.3. Cycle GAN

In natural environments, fruits are easily occluded by branches and leaves, which affects the accuracy of fruit size measurement. Existing occlusion processing methods are constrained by issues such as fruit shape or growth environments, making it difficult to balance accuracy and cost factors effectively. Therefore, an image restoration method is required to adaptively solve the fruit occlusion problem. This paper proposed a novel fruit-occlusion recovery method based on style transfers.

2.3.1. Generative adversarial networks

GAN (Generative adversarial networks) is a network model proposed by Goodfellow to generate data via an adversarial process, with a framework similar to that of a minimax two-player game (Goodfellow et al., 2020). GAN consist of two neural networks: a generator *G* and a

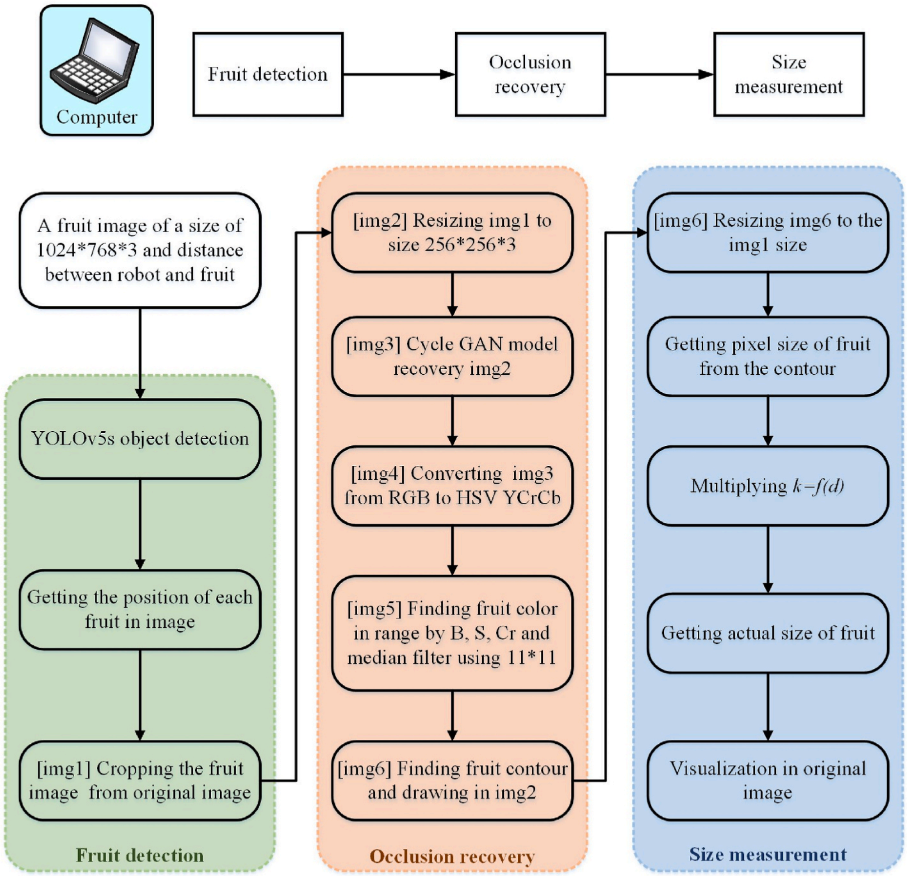
Computer
Fruit detection
Occlusion recovery
Size measurement
A fruit image of a size of 1024*768*3 and distance between robot and fruit
YOLOv5s object detection
Getting the position of each fruit in image
[img1] Cropping the fruit image from original image
Fruit detection
[img2] Resizing img1 to size 256*256*3
[img3] Cycle GAN model recovery img2
[img4] Converting img3 from RGB to HSV YCrCb
[img5] Finding fruit color in range by B, S, Cr and median filter using 11*11
[img6] Finding fruit contour and drawing in img2
Occlusion recovery
[img6] Resizing img6 to the img1 size
Getting pixel size of fruit from the contour
Multiplying k–f(d)
Getting actual size of fruit
Visualization in original image
Size measurement

Fig. 2. The specific size measurement process on the computer.

discriminator *D*. The generator generates fake data while the discriminator distinguishes the real data from the fake data. The noise vector *z* is used as an input to generate fake images similar to the real image *x*. The discriminator *D* is used to judge whether the given image is real or fake. An adversarial relationship exists between the generator *G* and discriminator *D*. In the GAN training process, *G* needs to make the generated sample closer to the real sample, that is, make *D(G(z))* close to 1. In contrast, *D* must be able to estimate the possibility that the sample comes from the real sample and the generated fake sample; that is, *D(x)* is closer to 1, and *D(G(z))* is closer to 0. Therefore, the target function of the GAN model is defined as follows (Goodfellow et al., 2020):

$$\min_{G}\max_{D} V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_{z}(z)}[\log(1-D(G(z)))] \tag{2}$$

Through such an adversarial process, the performances of the generator and discriminator are alternately improved. GAN have been applied in various domains, including image and video generation, style transfer, and data augmentation.

2.3.2. Cycle-consistent adversarial networks

The Cycle GAN is a neural network model expanded from the GAN (Zhu et al., 2017). It is composed of two mirrored GAN networks that form a ring network in the framework. They share two generators, *G* and *F*, each with discriminators D_Y and D_X. Each one-way GAN has two generators and one discriminator, as shown in Fig. 3. This structure makes its training independent of the paired images (Isola et al., 2017) and allows for better adaptive capabilities.

Here, *G* and *F* are generators that convert *X* real samples into *Y* samples and *Y* real samples into *X* samples, respectively. After the fake samples are generated, they are input into D_Y and D_X for identification, together with the real samples. This process corresponds to the adversarial loss functions (3) and (4), as inferred from Section 2.3.1:

$$L_{GAN}(G,D_Y) = E_{y\sim P_{data}(y)}[\log D_Y(y)] + E_{x\sim P_{data}(x)}[\log(1-D_Y(G(x)))] \tag{3}$$

$$L_{GAN}(F,D_X) = E_{x\sim P_{data}(x)}[\log D_X(x)] + E_{y\sim P_{data}(y)}[\log(1-D_X(G(y)))] \tag{4}$$

To make the generated image more similar to the expected image, the Cycle GAN introduces a cycle consistency loss function (5) (Zhu et al., 2017):

$$L_{cyc}(G,F) = E_{x\sim P_{data}(x)}\left[\|F(G(x))-x\|_1\right] + E_{y\sim P_{data}(y)}\left[\|G(F(y))-y\|_1\right] \tag{5}$$

Therefore, the loss function (6) includes two parts, as shown below:

$$L(G,F,D_X,D_Y) = L_{GAN}(G,D_Y) + L_{GAN}(F,D_X) + \lambda L_{cyc}(G,F) \tag{6}$$

where λ is used to control the relative importance of the two parts of the loss.

The target function of the Cycle GAN network model is defined as (Zhu et al., 2017):

$$G^*,F^* = \arg\min_{G,F}\max_{D_X,D_Y} L(G,F,D_X,D_Y) \tag{7}$$

With continuous improvement in the alternate training performance, the two generators can finally realize the mutual generation of *X* and *Y* style images. The unsupervised and adaptive approach of Cycle GAN for style transfer between two images provides a powerful reference for its application in food engineering. For instance, GANana achieved two-to-three-dimensional image reconstruction of isolated banana (Hartley et al., 2021). Fruit occlusion is the main factor affecting fruit size measurement. Based on the idea of Cycle GAN style transfer, occluded and un-occluded images of fruit were input as two different style images into the Cycle GAN model for training. The occluded style could be used to generate an un-occluded style image to achieve fruit occlusion recovery. Compared with existing occlusion processing methods, this method is not limited by the fruit contour and is not easily disturbed by environmental factors.

2.4. Dataset acquisition

2.4.1. YOLOv5 training dataset

To obtain highly accurate object detection, a significant number of effective images are often required for training. The acquisition of a dataset can be established by oneself or from a public dataset. Self-built datasets contain specific applicable objects that are typically collected from the field and labeled manually. As shown in Fig. 4a, fruit images of two categories, citrus and strawberry, were collected in natural environments. Citrus images were primarily captured in citrus orchard, and strawberry images were obtained from greenhouse. The self-built

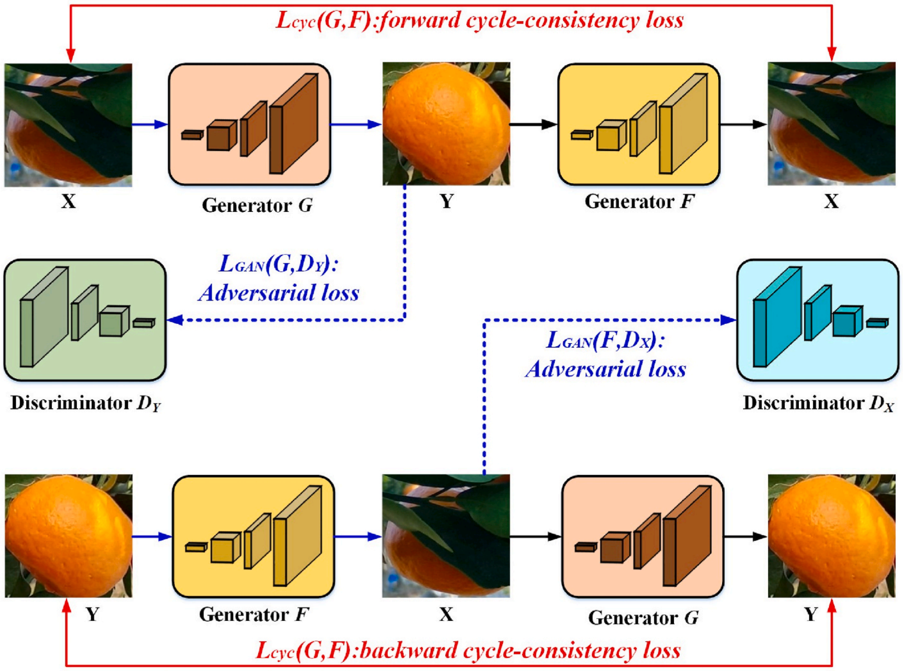

Fig. 3. The overall framework of the Cycle GAN, including two adversarial loss, forward cycle-consistency loss and backward cycle-consistency loss.

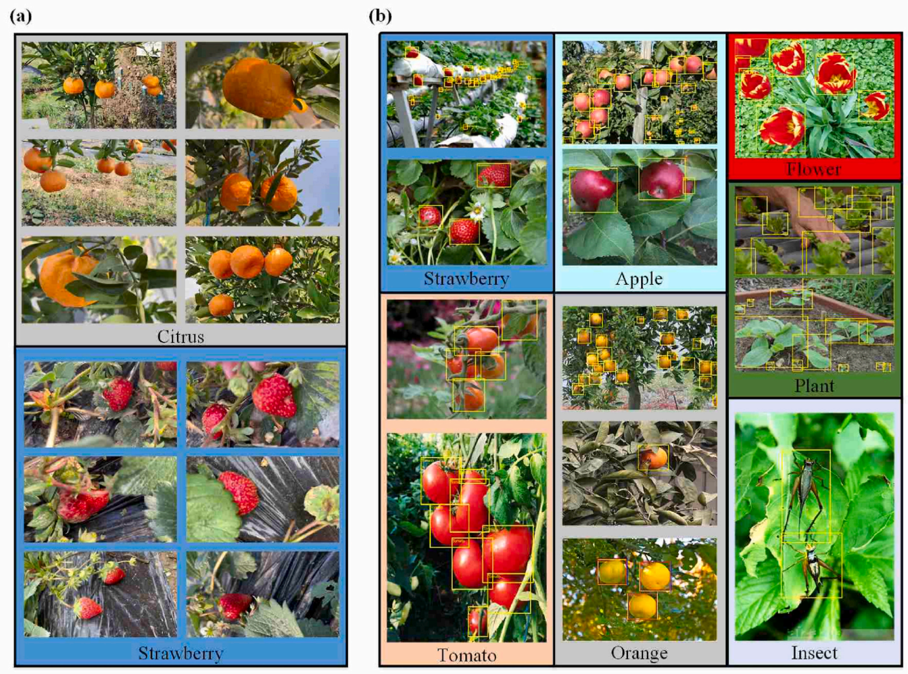

Fig. 4. (a) Self-built datasets of citrus and strawberry; (b) Some fruit, flower, plant and insect images obtained from Open Images website Open Images V4 (storage.googleapis.com).

dataset includes images of un-occluded fruits, fruits occluded by leaves, branches, overlapping fruits, and fruits affected by lighting conditions. A camera (Nikon) was used as the capturing tool, and the image size was set at 1920*1080*3, ensuring that the fruits are clearly and completely displayed in the images. However, the workload of manual collection is impractical and the image richness is usually insufficient; therefore, public datasets need to be used to supplement private datasets.

Public image datasets have been increasingly used in agriculture in recent years, including datasets specifically for fruits, such as date fruit (Altaheri et al., 2019), deep fruit (Sa et al., 2016), Open Images (Kuznetsova et al., 2020) and COCO(Lin et al., 2014). The chosen image dataset for the training and validation was from the Open Images dataset (Kuznetsova et al., 2020), which contains nine million images and label data files corresponding to some of the images. More than 600 object classes of tags can be directly used for supervised learning, which greatly reduces the workload of image annotation in dataset preparation. The available detection categories for agricultural application scenarios include fruits (apples, oranges, strawberries, tomatoes, cucumbers, etc.), flowers, plants, and insects. The detection objects are framed in a box, as shown in Fig. 4b. As shown in Table 1, after image collection and screening, a total of 1213 citrus images were obtained, including 667 images from the self-built database and 546 images from the Open Images dataset. A total of 1387 strawberry images were collected, including 714 from the self-built database and 613 from the Open Images dataset. The dataset contains a total of 1381 self-built images and 1219 Open Images, with a ratio of approximately 1.13, indicating a relatively balanced distribution. Combining two datasets can increase the diversity of training samples, which helps to improve the generalization ability and robustness of the model.

Table 1
The number of YOLOv5 training dataset.

Class	Self-built dataset	Open Images	Self-built: Open Images
Citrus	667	546	1.22
Strawberry	714	673	1.06

Table 2
The number of Cycle GAN training dataset.

Class	Fruit occluded		Fruit un-occluded	
	Manual	Natural	Manual	Natural
Citrus	4075	7188	3927	3987
Strawberry	3947	4515	1288	4923

2.4.2. Cycle GAN training dataset

According to the analysis in Section 2.3, the Cycle GAN training dataset requires images of a certain class of occluded and un-occluded. Few existing public datasets meet these requirements, and for the purpose of this study, a training dataset needed to be built around these requirements. Taking citrus as an example, the dataset preparation process is shown in Fig. 5.

(1) Due to the large number of training images required by Cycle GAN, a method that decomposes videos into images is considered to quickly obtain original images. In order to obtain diverse citrus occluded and un-occluded images, citrus videos are collected from both manual setting and natural environments. For the manual setting, individual citrus fruit was recorded from various angles and positions, with occlusions including leaves and branches at different angles and proportions, under a single background condition. In the natural environment, citrus videos are captured from different angles and positions, presenting more complex challenges such as occlusions caused by leaves and branches, fruit-to-fruit occlusions, and lighting variations. By combining data from these two distinct environments, the richness of the training dataset can be significantly enhanced, contributing to the development of more robust models. The capturing tool used here is a Nikon digital camera, and the extracted image size is 1920*1080*3. The video frame rate is set at 30 frames per second, with 5 images extracted per second.

(2) Citrus images undergo operations such as grayscale conversion and binarization, enabling the extraction of citrus contour region based on contour size and shape feature.

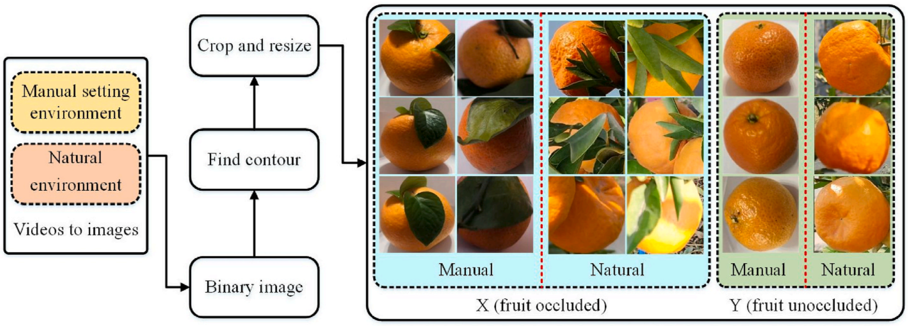

Fig. 5. The overall flow-chart of Cycle GAN dataset preparation.

(3) After the edge coordinates of the fruit contour are detected, an image with only the part of the fruit and branches and leaves is obtained through a cropping operation. The cropped image is normalized in size and converted into a size of 256*256*3. As shown in Table 2, a total of 11,263 occluded citrus images and 7914 un-occluded citrus images were obtained, along with 8462 occluded strawberry images and 6211 un-occluded strawberry images. Among them, 13237 images were acquired in manual setting environments, while 20613 images were captured in natural environments. This dataset contains as much as possible various occluded and un-occluded situations that may be encountered in practical applications.

2.5. Image processing

After the object detection model is processed, a single-fruit image is cropped. An image processing flowchart is shown in Fig. 6a. For the fruit size information, it can be obtained by combining the distance values. The image processing programs used in this study were developed using Python OpenCV.

(1) The RGB image of the fruit is read and input to the trained Cycle GAN network for occlusion recovery.
(2) To more accurately extract the contour of fruits from recovered image, conversion from RGB to HSV and YCrCb color spaces was performed and split into single-channel images. As shown in Fig. 7, the B, S, and Cr channels are selected as they exhibited more distinct fruit contour, where S and Cr respectively describe saturation and chrominance.

(3) The B, S, Cr variation range of a specific color is then used to extract the fruit image mask:

$$\text{mask} = \begin{cases} 255, (B,S,Cr)_{(i,j)} \in \left[(B,S,Cr)_{color_{low}}, (B,S,Cr)_{color_{high}}\right] \\ 0, (B,S,Cr)_{(i,j)} \notin \left[(B,S,Cr)_{color_{low}}, (B,S,Cr)_{color_{high}}\right] \end{cases} \tag{8}$$

In the equation above, $(B,S,Cr)_{(i,j)}$ represents the B,S,Cr pixel value corresponding to the position *(i, j)*. $(B,S,Cr)_{color\_low}$ and $(B,S,Cr)_{color\_high}$ represent the lower and upper limits, respectively, where the B, S, Cr pixel value within this range corresponding to a specific color.

(4) The median filtering operation filters the isolated noise points, extracts the contour of the fruit, and covers the original image. As shown in Fig. 6b, the pixel size, including the length, width, and area, can be calculated from the contour.
(5) The ratio coefficient *k* between the actual size and the pixel size is obtained by combining the distance *d.* After conducting multiple tests within the range of 10–50 cm between the camera and the

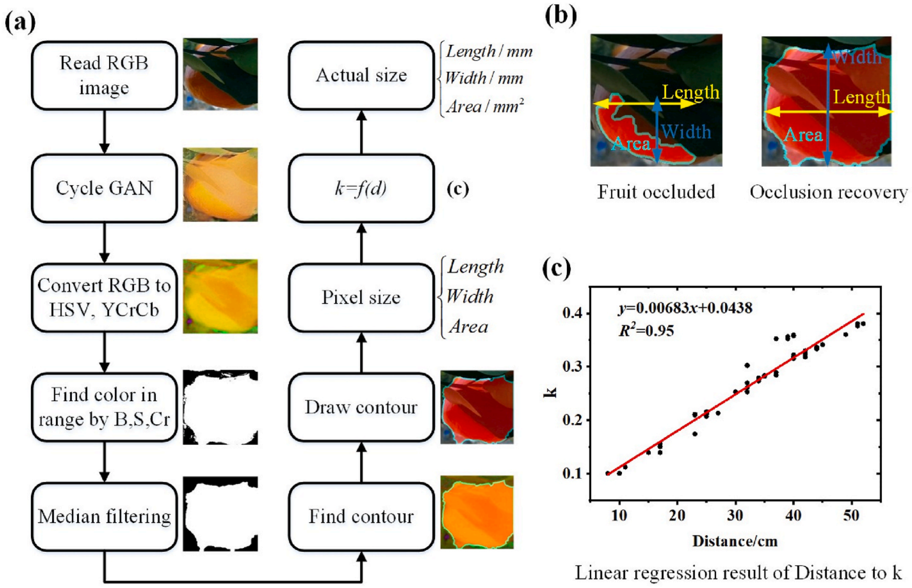

Fig. 6. Image processing for size measurement. (a)The overall flow of image processing; (b)The size measurement of fruit occluded and un-occluded; (c)The linear regression result of Distance to *k*.

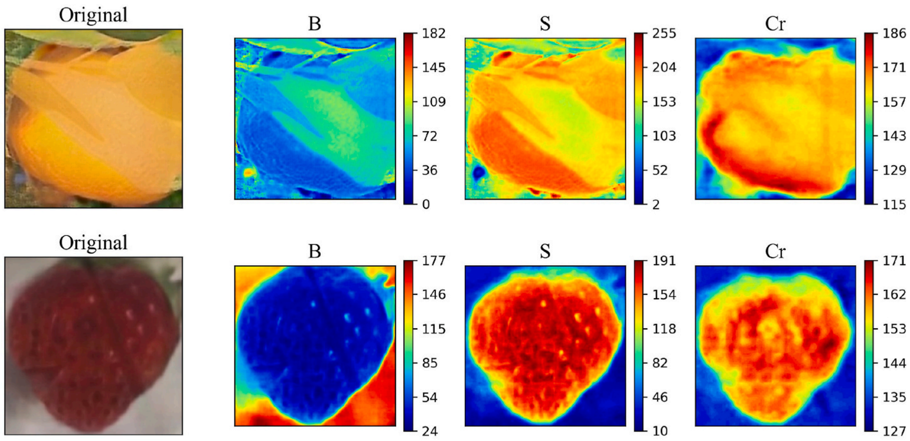

Fig. 7. The original color fruit recovered images and corresponding B, S, Cr single-channel images.

plant, the relationship diagram is obtained by linear regression, as shown in Fig. 6c.

$$k = 0.00683 * d + 0.0438 \; R^2 = 0.95 \tag{9}$$

(6) The pixel size obtained in (4) is multiplied by the ratio *k* of (5) to obtain the actual size.

The actual area is measured using Equation (10):

$$\text{area}(actual_{size}) = \frac{area_{contour}}{area_{box}} * Length * Width(actual_{size}) \tag{10}$$

where $area_{contour}$ and $area_{box}$ represents the fruit contour pixel area and the detected rectangular box pixel area, respectively, and *Length* and *Width* are the actual sizes.

2.6. Performance metrics

In order to assess the algorithm performance in this study, a series of metrics have been defined for the stage of object detection, occlusion recovery, and size measurement. The following is an introduction to the evaluation metrics used for each stage.

2.6.1. Object detection

Precision, Recall, mAP_{50}, $\text{mAP}_{50:90}$ and model size are the five metrics used to evaluate the performance of the object detection model.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

Precision measures the proportion of correctly predicted positive instances out of all positive predictions, while Recall measures the proportion of correctly predicted positive instances out of all actual positive instances. TP represents true positives, which are the correctly predicted positive instances. FP represents false positives, which are the incorrectly predicted positive instances. FN represents false negatives. mAP_{50} refers to the mean Average Precision when the intersection over union (IOU) is 0.5, while $\text{mAP}_{50:90}$ computes the mean average precision across a range of IOU thresholds from 0.5 to 0.95 at intervals of 0.05. mAP_{50} and $\text{mAP}_{50:90}$ are an important evaluation metric that provides an overall performance measure for object detection models. The closer Precision, Recall, mAP_{50} and $\text{mAP}_{50:90}$ are to 1, the better the performance of the object detection model. Model size is often considered as a reference for evaluating the running speed and lightweight characteristics of a model.

2.6.2. Occlusion recovery

In order to evaluate the performance of occlusion recovery, evaluation metrics such as relative error of contour pixel area, intersection over union of contour ($\text{IOU}_{\text{contour}}$), under-segmentation rate (UR) and over-segmentation rate(OR) were employed.

$$\text{Relative error} = \frac{|R_s - T_s|}{R_s} \times 100\% \tag{13}$$

$$IOU_{contour} = \frac{R_s \cap T_s}{R_s \cup T_s} \times 100\% = \frac{I_s}{R_s + O_s} \times 100\% \tag{14}$$

$$UR = \frac{U_s}{R_s + O_s} \times 100\% \tag{15}$$

$$OR = \frac{O_s}{R_s + O_s} \times 100\% \tag{16}$$

As shown in Fig. 8, Rs represents the contour of the un-occluded fruit, which serves as the reference result. Ts represents the fruit contour restored by Cycle GAN after being occluded, which is the occlusion recovery result. The relative error between Rs and Ts is used to evaluate the numerical deviation. $\text{IOU}_{\text{contour}}$ represents the proportion of correctly recovered occlusions Is to the union of Rs and Ts. Us represents the area that appears in the reference result but not in the occlusion recovery result, while Os represents the area that appears in the occlusion recovery result but not in the reference result. $\text{IOU}_{\text{contour}}$, UR, and OR are calculated at the pixel level to assess the overlap between the occlusion recovery result and the reference result.

2.6.3. Size measurement

The evaluation metrics for size measurement include the mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE) between the measured results and the actual results. The size parameters involved in the evaluation are distance,

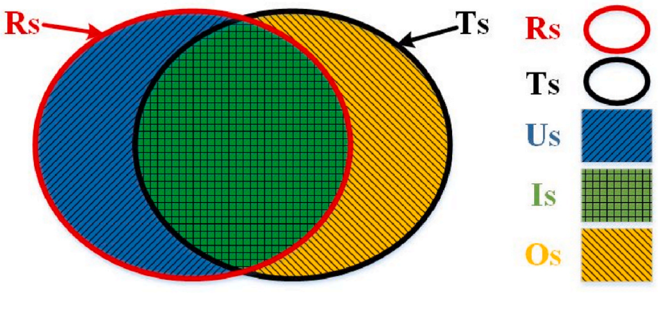

Fig. 8. Schematic diagram of occlusion recovery evaluation metrics definition.

length, width, and area.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \tag{17}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{18}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{19}$$

3. Results and discussion

In order to systematically evaluate the algorithm performance of this study, the test data involves two datasets. The validation results for object detection performance in Section 3.1 are based on the object detection training dataset. Other results, including the object detection accuracy comparison in Section 3.1, as well as all the test results in Sections 3.2, 3.3, and 3.4, are based on the same field fruit images collected using the robotic platform. The image size is 1024*768*3. These field fruit images comprise 99 citrus images and 57 strawberry images from the natural environment, with a total of 161 oranges and 66 strawberries. Among them, there are 48 pairs of citrus and 28 pairs of strawberries that were captured under both occluded and un-occluded conditions, specifically intended for evaluating occlusion recovery performance.

3.1. The performance on fruit detection

The 1213 citrus images and 1387 strawberry images from the training dataset were divided into a training set and a validation set according to the ratio of 8:2, respectively. To compare the object detection performance of YOLOv5, Faster RCNN and SSD were selected for comparative experiments using the same training set and validation set. The stable model obtained after 300 epochs were chosen as the final model. The performance comparison of the three object detection models is shown in Table 3. YOLOv5 achieved the best results for the same fruit in terms of various accuracy metrics. The average values of Precision, Recall and mAP_{50} were all above 0.8. Faster RCNN performed slightly lower than YOLOv5, with mAP_{50} reaching a similar level to YOLOv5. SSD showed relatively general performance. Additionally, YOLOv5s had the smallest model size (14M) compared to the Faster RCNN(51M) and SSD(22M), which benefits both running speed and model lightweight.

To further evaluate the generalization ability and robustness of the models, fruit images collected from the robotic platform were used as the test set to observe the detection accuracy of the three object detection models on a new dataset. As shown in Table 4, Faster R–CNN achieved a 96.04% accuracy with its two-stage object detector processing method. YOLOv5s, on the other hand, achieved a 95.60% high-precision detection. Lastly, SSD achieved an accuracy rate of 75.33%. In terms of time consumption, the average processing time of YOLOv5s per image was only 0.05 s. This is significantly lower than those of Faster RCNN and SSD, whose average processing time are 2.7 s and 1.4 s respectively. This can be attributed to the fact that YOLOv5s has the smallest model size. YOLOv5s achieves object detection accuracy comparable to Faster RCNN in a shorter period of time, making it the best choice in terms of overall performance.

3.2. The performance on occlusion recovery

A total of 11263 occluded citrus images and 7914 un-occluded citrus images, as well as 8462 occluded strawberry images and 6211 un-occluded strawberry images, were used as the Cycle GAN training dataset. After 8 h of training, the total loss value stabilized and was used as the final model. To calculate the occlusion recovery performance metrics of Cycle GAN, 48 pairs of oranges and 28 pairs of strawberries were tested using fruit images collected by the robotic platform. The test results for citrus and strawberries are shown in Fig. 9 and Fig. 10, respectively. Figs. 9a and 10a show the fruit occlusion recovery images, where.

(1) the un-occluded fruit image;
(2) the same fruit as in (1) under natural occlusion;
(3) the contour of fruit (1), which serves as the reference result Rs;
(4) the contour of the fruit (2) after occlusion recovery using Cycle GAN, which serves as the occlusion recovery result Ts;
(5) the occlusion recovery contour obtained using the traditional Hough circle detection algorithm for citrus fruits.

To ensure a uniform evaluation, the $\mathrm{IOU}_{\mathrm{contour}}$ is calculated using the value 100 - $\mathrm{IOU}_{\mathrm{contour}}$. The closer the relative error, 100- $\mathrm{IOU}_{\mathrm{contour}}$, UR, and OR are to 0, the better the occlusion recovery performance. As shown in Fig. 9b, for citrus fruits, Cycle GAN exhibits a smaller range of variations in all four metrics compared to the Hough circle detection algorithm, indicating stable performance. The specific data distributions for each metric are shown in Fig. 9c-f. The evaluation metric values for Cycle GAN remain below 20%, while some of the data points for Hough reach above 80%. For strawberry, the conventional circle-like fitting method cannot be applied for occlusion recovery due to the irregular shape. As shown in Fig. 10b-f, the occlusion recovery evaluation metrics of Cycle GAN also exhibit a relatively stable distribution within 20%. The mean values of each occlusion recovery evaluation metrics are shown in Table 5. The mean values of Cycle GAN evaluation metrics relative error, 100- $\mathrm{IOU}_{\mathrm{contour}}$, UR, and OR are approximately 5.48%, 11.19%, 6.68%, and 5.01%, which are much lower than those of the Hough method. Cycle GAN achieves high accuracy in all evaluation metrics, effectively fulfilling occlusion recovery under natural conditions.

In addition, Table 6 presents the comparison results between Cycle GAN and other fruit occlusion recovery algorithms(Ge et al., 2019; Gong et al., 2022; Kim et al., 2023; Magistri et al., 2022). It can be observed that Cycle GAN achieved good recovery performance. (Gong et al., 2022) achieved low recovery error due to the use of high-precision RGB-D cameras. The effective recovery of occluded areas can be achieved by combining multi-modal information such as fruit images and corresponding depth maps, however this also results in increased costs.

3.3. The performance on size measurement

After obtaining the fruit contours and distance information, the fruit

Table 3
The performance comparison of the three object detection models.

Model	Class	Precision	Recall	mAP50	mAP50:95	Model size/M
Faster RCNN	Citrus	0.72	0.67	0.84	0.62	51
	Strawberry	0.57	0.54	0.77	0.49	51
SSD	Citrus	0.31	0.30	0.38	0.25	22
	Strawberry	0.19	0.23	0.26	0.16	22
YOLOv5s (this work)	Citrus	0.88	0.83	0.88	0.72	14
	Strawberry	0.79	0.74	0.82	0.57	14

Table 4
The comparison of object detection accuracy on fruit images collected from the robotic platform.

Model	Class	number	Correctly detected	Falsely detected	Accuracy/%	Time consumption/s
Faster RCNN	Citrus	161	157	4	96.04	2.72
	Strawberry	66	61	5		
SSD	Citrus	161	114	42	75.33	1.4
	Strawberry	66	57	9		
YOLOv5s (this work)	Citrus	161	155	6	95.60	0.05
	Strawberry	66	62	4		

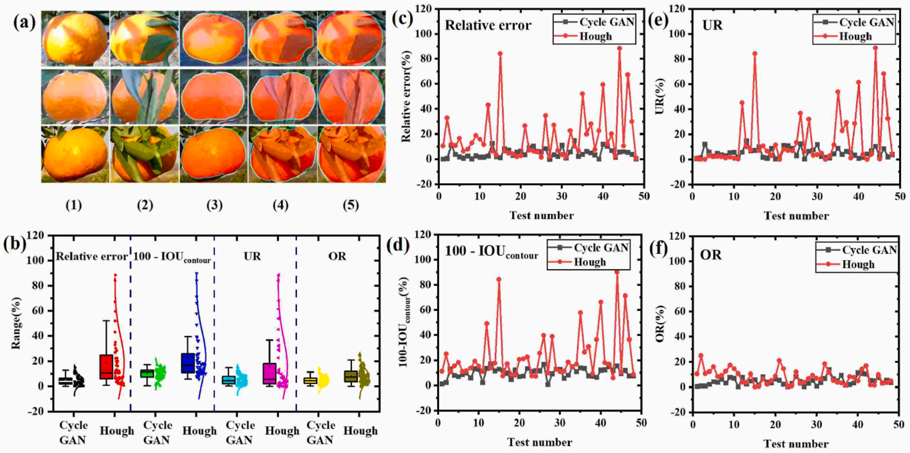

Fig. 9. Occlusion recovery using Cycle GAN. (a) Comparison of occlusion recovery of citrus; (b) Box plot of the relative error, 100- IOU$_{contour}$, UR, and OR of the Cycle GAN and Hough; The (c) relative error, (d)100- IOU$_{contour}$, (e)UR, and (f)OR occlusion recovery of 48 pairs of citrus.

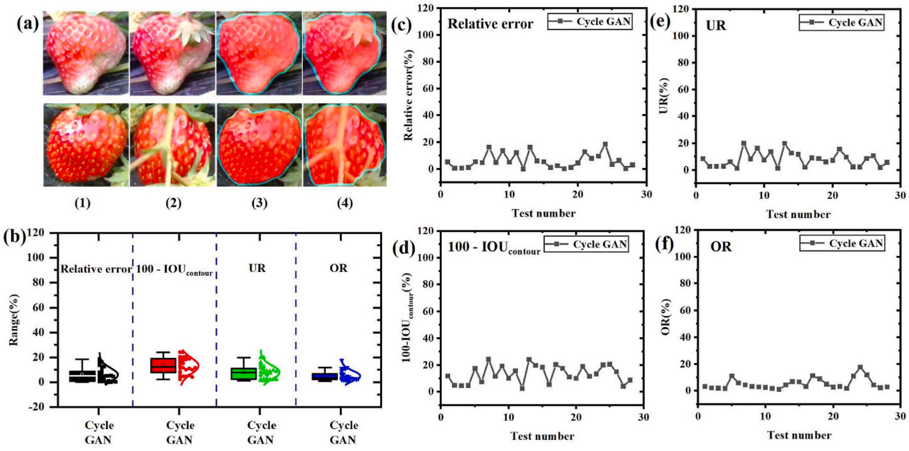

Fig. 10. Occlusion recovery using Cycle GAN. (a) Comparison of occlusion recovery of strawberry; (b) Box plot of the relative error, 100- IOUcontour, UR, and OR of the Cycle GAN; The (c) relative error, (d)100- IOUcontour, (e)UR, and (f)OR occlusion recovery of 28 pairs of strawberry.

Table 5
The mean values of occlusion recovery evaluation metrics.

Method	Class	Relative error/%	100-IOUcontour/%	UR/%	OR/%
Cycle GAN (this work)	Citrus	4.83	10.07	5.28	4.78
	Strawberry	6.12	12.31	8.08	5.23
Hough	Citrus	18.76	23.53	15.33	8.21

size was measured according to the process shown in Fig. 6a. To test the accuracy of the size measurements, size measurement evaluation metrics were computed for 155 citrus fruits and 62 strawberries in terms of the comparison between their actual sizes and the measured sizes based on the same field fruit images collected using the robotic platform. The actual size was measured as shown in Fig. 11. The length and width were measured with Vernier calipers while the distance was determined using a ruler. The fruit area was calculated from $area_{contour}$ and $area_{box}$, which represents the fruit contour pixel area and the rectangular box pixel area, respectively.

Fig. 12 presents the box plots of the actual and measured values for distance, length, width, and area of citrus and strawberry. It is evident that the measured values closely align with the actual values in terms of their numerical distribution. To further analyze the errors in each size, corresponding MAPE, RMSE, and MAE were calculated as shown in Table 7. The distance measurement MAPE can be better maintained at 10%, which provides a strong guarantee of the subsequent size measurement accuracy. For the length and width measurements, the MAPE, RMSE, and MAE are approximately 9.18%, 5.38 mm and 4.24 mm. A fast

Table 6
Comparison of with other algorithms for fruit occlusion recovery.

Algorithm	Class	Recovery accuracy/%	Recovery error/%	Deep learning
YOLOv5s + Cycle GAN (this work)	Citrus/ Strawberry	–	5.48	yes
YOLOv5s + Hough circle	Citrus	–	18.76	no
DCNN + WHR (Ge et al., 2019)	Strawberry	87.00	–	yes
Mask RCNN + SPR (Gong et al., 2022)	Tomato	–	2.15	yes
U-net (Kim et al., 2023)	Cucumber	82.43	–	yes
FCN + encoder (Magistri et al., 2022)	Strawberry	87.97	–	yes

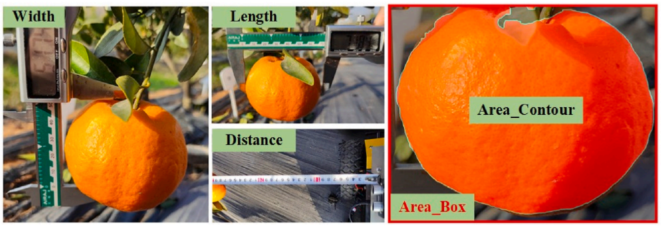

Fig. 11. The measurement operations of width, length distance and actual area.

and low-cost measurement is achieved using a noncontact measurement of the basic size in the complex natural environment. For the area, the MAPE, RMSE, and MAE of the measurement was 14.23%, 353.98 mm^2 and 277.45 mm^2, which is lower to the relative error levels of area measurement in existing reports (22%, 20%) (Golbach et al., 2016; Masuda, 2021). However, the methods employed in these reports utilize multi-camera, multi-angle imaging to reconstruct plants in non-natural environment, which is applicable only in fixed scenarios and involves high costs.

3.4. Field test performance and discussion

Illumination and occlusion under natural conditions pose challenges to noncontact fruit size measurements. Hence, this study proposes a size measurement method based on YOLOv5 and Cycle GAN. However, in practical applications, issues such as time, cost, and environmental adaptability must be considered. Therefore, field tests were conducted to test the practicality. Following the procedure shown in Fig. 1, fruit images and distance values were collected in a natural environment and wirelessly transmitted to a computer. A total of 99 citrus images (161 citrus) and 57 strawberry images (66 strawberries) were collected, and the real situations including the occlusion caused by leaves, branches, and fruits, as well as lighting problem. The object detection, occlusion recovery, and size measurement test results based on these fruits image have been shown in section3.1, 3.2 and 3.3. The final field test results are presented in Figs. 13 and 14, after rapid processing of the collected images, the information such as the class, length, width and area of the fruit is clearly visualized in the original image. The corresponding results are stored in the computer in the form of images and data, which is convenient for subsequent analysis and management. These results indicate the proposed method can effectively adapt to the size measurement in the natural environment under different occlusion and illumination conditions. However, this method is not applicable when the fruit is completely obscured. Other natural conditions, such as blurring and overexposure, require further testing.

For the overall test results, the average measurement time was 0.2 s while the detection accuracy was 95.6%. The average relative error of

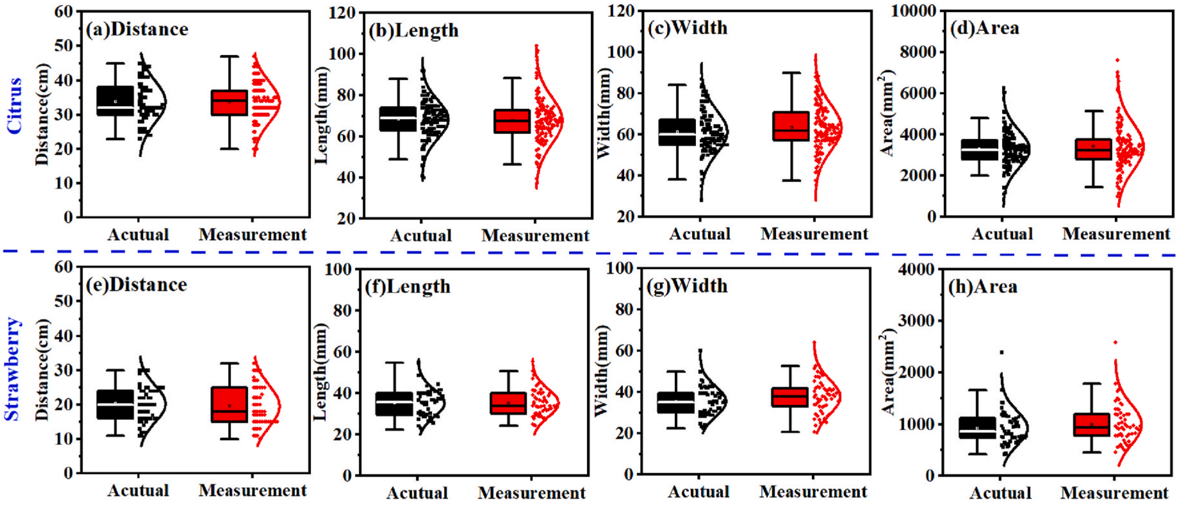

Fig. 12. The citrus actual and measurement result of (a) distance, (b) length, (c) width, and (d) area. The strawberry actual and measurement result of (a) distance, (b) length, (c) width, and (d) area.

Table 7
The MAPE, RMSE, and MAE for size measurement.

Class	Distance			Length			Width			Area		
	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MRPE	RMSE	MAE	MRPE	RMSE	MAE
Citrus	6.76	2.83	2.23	6.56	6.21	4.58	8.20	6.31	4.91	12.24	547.64	414.32
Strawberry	9.05	2.34	1.82	10.91	4.53	3.70	11.03	4.46	3.75	16.22	160.33	140.59
Average	7.90	2.58	2.03	8.73	5.37	4.14	9.62	5.39	4.33	14.23	353.98	277.45
unit	%	cm	cm	%	mm	mm	%	mm	mm	%	mm^2	mm^2

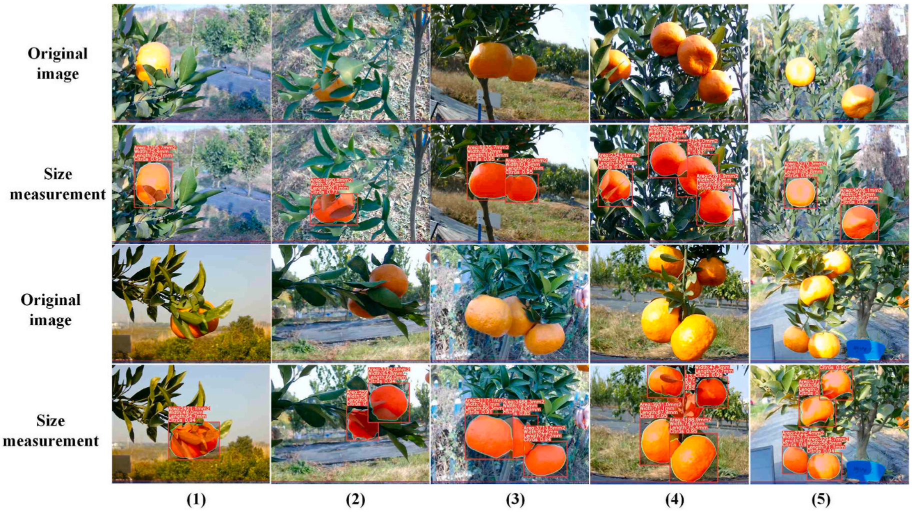

Fig. 13. Field test results of citrus. Occlusion caused by (1) leaves, (2) leaves and branches, (3) fruits, (4) leaves and fruits, and (5) lighting problem.

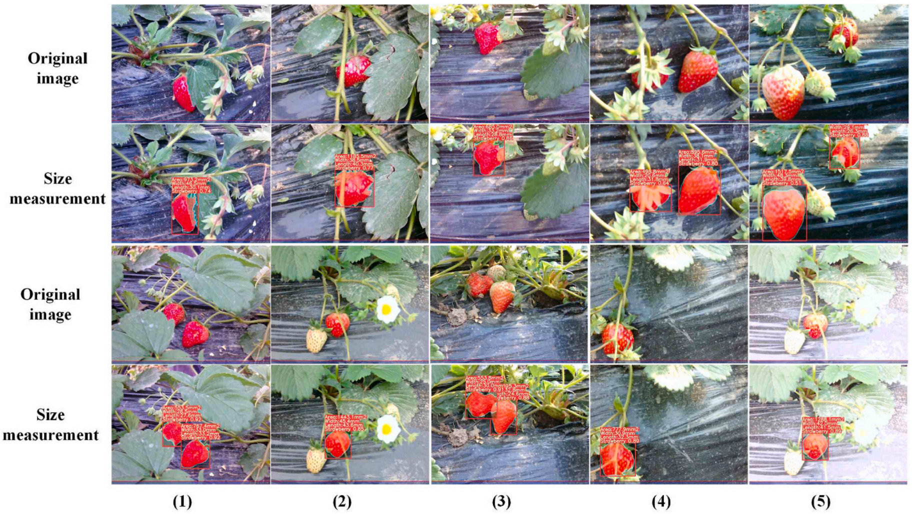

Fig. 14. Field test results of strawberry. Occlusion caused by (1) leaves, (2) branches, (3) fruit, (4) leaves and branches, and (5) lighting problem.

the occlusion recovery was 5.48%. It is worth noting that the average relative error in the measurement of individual fruit sizes is 8.43% for individual fruits and the 10.12% overall error under the complex natural environment. However, in the case of multiple fruits, the measurement errors are larger due to issues such as the presence of branches and leaves, mutual occlusion between fruits, and varying distances. In comparison to existing solutions that utilize depth cameras to achieve high-precision measurements of individual fruits, obtaining field test results using our method in complex natural environments with low-cost ranging devices is deemed acceptable. In terms of speed and accuracy, the proposed method achieved a good usage effect.

Although the proposed method realizes satisfactory fruit size measurement results in natural environments, there are still some limitations. On one hand, the image and distance data require collecting from a fixed distance and angle as much as possible. In this study, a data collection box is used to assist in positioning, and other appropriate method can be selected according to the requirements in future. On the other hand, it should be noted that the input images of Cycle GAN should ensure that the citrus is located in the central area.

4. Conclusion

To achieve a noncontact measurement of fruit size in the natural growth state, this study proposed a systematic fruit size measurement method based on YOLOv5 and Cycle GAN. Additionally, it built a robot platform that automatically collects fruit images and distance information at close range and transmits it to a computer for real-time object detection, occlusion recovery, and size measurement. The object detection algorithm based on YOLOv5 showed better generalization ability than other comparable methods and achieved recognition accuracy of 95.6% in field tests. Occlusion recovery based on Cycle GAN adaptively eliminated the problem of being occluded by branches, leaves and fruits in the fruit image and was not limited by the shape of the fruit. In addition, the average relative error was 5.48%, which satisfies the actual measurement requirements. Through low-cost distance measurement, a proportional relationship between the actual size and pixel size can be easily obtained. Additionally, noncontact fruit size measurement can be achieved with an average relative error of approximately 8.43% for individual fruits and the 10.12% overall error under the complex natural environment and a single image processing time of only 0.2 s. In practice, the entire system ran in real time with multiple threads, which meets the requirements for rapid detection in large batches.

The scheme proposed in this study can be improved in practical operations. Primarily, the measurement component can migrate from the computer terminal to the embedded device; therefore, it is necessary to explore auxiliary measurement and positioning solutions.

Funding

This work was supported by Independent Innovation Fund of Huazhong University of Science and Technology.

CRediT authorship contribution statement

Bingkai Wang: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Mengqi Li:** Visualization, Validation, Investigation, Data curation. **Yuqi Wang:** Validation, Investigation, Data curation. **Yuhan Li:** Validation, Investigation, Data curation. **Zilan Xiong:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Al-kaf, Hasan Ali Gamal, Chia, Kim Seng, Mubin, Faiz Ridwan Bin Abdul, 2020. The size and weight prediction for intact pineapples using A low cost vision system. In: 2020 Zooming Innovation in Consumer Technologies Conference (ZINC). IEEE, 146–50. https://ieeexplore.ieee.org/document/9161770/.

Altaheri, Hamdi, et al., 2019. Date fruit dataset for intelligent harvesting. Data Brief 26, 104514. https://doi.org/10.1016/j.dib.2019.104514.

Bochkovskiy, Alexey, Wang, Chien-Yao, Liao, Hong-Yuan Mark, 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. http://arxiv.org/abs/2004.10934.

Chen, Jiqing, et al., 2021. Detecting ripe fruits under natural occlusion and illumination conditions. Comput. Electron. Agric. 190 (September), 106450 https://doi.org/10.1016/j.compag.2021.106450.

Donmez, Cenk, Villi, Osman, Berberoglu, Suha, Cilek, Ahmet, 2021. Computer vision-based citrus tree detection in a cultivated environment using UAV imagery. Comput. Electron. Agric. 187 (February), 106273 https://doi.org/10.1016/j.compag.2021.106273.

Fan, Shuxiang, et al., 2020. On line detection of defective apples using computer vision system combined with deep learning methods. J. Food Eng. 286 (November 2019), 110102 https://doi.org/10.1016/j.jfoodeng.2020.110102.

Feng, Yu, Li, Xinxing, Zhang, Yinggang, Xie, Tianhua, 2023. Detection of atlantic salmon residues based on computer vision. J. Food Eng. 358 (January), 111658 https://doi.org/10.1016/j.jfoodeng.2023.111658.

Gan, Hao, Lee, Won S., Alchanatis, Victor, Abd-Elrahman, A., 2020. Active thermal imaging for immature citrus fruit detection. Biosyst. Eng. 198, 291–303. https://doi.org/10.1016/j.biosystemseng.2020.08.015.

Ge, Yuanyue, Xiong, Ya, Pål, J., 2019. Instance segmentation and localization of strawberries in farm conditions for automatic fruit harvesting. From IFAC-PapersOnLine 52 (30), 294–299.

Golbach, Franck, et al., 2016. Validation of plant Part Measurements using a 3D reconstruction method suitable for high-throughput seedling phenotyping. Mach. Vis. Appl. 27 (5), 663–680.

Gong, Liang, Wang, Wenjie, Wang, Tao, Liu, Chengliang, 2022. Robotic harvesting of the occluded fruits with a precise shape and position reconstruction approach. J. Field Robot. 39 (1), 69–84.

Goodfellow, Ian, et al., 2020. Generative adversarial networks. Commun. ACM 63 (11), 139–144. https://dl.acm.org/doi/10.1145/3422622.

Guo, Zhiming, et al., 2020. Quantitative detection of apple watercore and soluble solids content by near infrared transmittance spectroscopy. J. Food Eng. 279 (September 2019), 109955 https://doi.org/10.1016/j.jfoodeng.2020.109955.

Hartley, Zane K.J., Jackson, Aaron S., Pound, Michael, French, Andrew P., 2021. Ganana: unsupervised domain adaptation for volumetric regression of fruit. Plant Phenomics 2021.

Hmidani, O., Ismaili Alaoui, E.M., 2022. A comprehensive survey of the R-CNN family for object detection. In: Proceedings - 2022 5th International Conference on Advanced Communication Technologies and Networking, CommNet, pp. 1–6, 2022.

Isola, Phillip, Zhu, Jun Yan, Zhou, Tinghui, Efros, Alexei A., 2017. Image-to-Image translation with conditional adversarial networks. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-Janua, pp. 5967–5976.

Jadhav, Tushar, Singh, Kulbir, Abhyankar, Aditya, 2019. Volumetric estimation using 3D reconstruction method for grading of fruits. Multimed. Tool. Appl. 78 (2), 1613–1634. http://link.springer.com/10.1007/s11042-018-6271-3.

Kim, Sungjay, et al., 2023. Application of amodal segmentation on cucumber segmentation and occlusion recovery. Comput. Electron. Agric. 210 (April), 107847 https://doi.org/10.1016/j.compag.2023.107847.

Kuznetsova, Alina, et al., 2020. The open images dataset V4: unified image classification, object detection, and visual relationship detection at scale. Int. J. Comput. Vis. 128 (7), 1956–1981. https://doi.org/10.1007/s11263-020-01316-z.

Lin, Tsung Yi, et al., 2014. Microsoft COCO: common objects in context. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8693 LNCS(PART 5), 740–55.

Liong, Sze Teng, Wu, Yi Liang, Liong, Gen Bing, Gan, Y.S., 2023. Moving towards agriculture 4.0: an AI-AOI carrot inspection system with accurate geometric properties. J. Food Eng. 357 (January), 111632 https://doi.org/10.1016/j.jfoodeng.2023.111632.

Lu, Shenglian, Chen, Wenkang, Zhang, Xin, Karkee, Manoj, 2022. Canopy-Attention-YOLOv4-Based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. Comput. Electron. Agric. 193 (January), 106696 https://doi.org/10.1016/j.compag.2022.106696.

Magistri, Federico, et al., 2022. Contrastive 3D shape completion and reconstruction for agricultural robots using RGB-D frames. IEEE Rob. Autom. Lett. 7 (4), 10120–10127.

Masuda, Takeshi, 2021. Leaf area estimation by semantic segmentation of point cloud of tomato plants. In: Proceedings of the IEEE International Conference on Computer Vision, 2021-October: 1381–89.

Miranda, Juan C., et al., 2023. Fruit sizing using ai: a review of methods and challenges. Postharvest Biol. Technol. 206 (February).

Phate, Vikas R., Malmathanraj, R., Palanisamy, P., 2021. Classification and indirect weighing of sweet lime fruit through machine learning and meta-heuristic approach. Int. J. Fruit Sci. 21 (1), 528–545. https://doi.org/10.1080/15538362.2021.1911745.

Redmon, Joseph, Ali, Farhadi, 2018. YOLOv3: an Incremental Improvement. http://arxiv.org/abs/1804.02767.

Sa, Inkyu, et al., 2016. Deepfruits: a fruit detection system using deep neural networks. Sensors 16 (8).

Sari, Yuita Arum, Gofuku, Akio, 2023. Measuring food volume from RGB-depth image with point cloud conversion method using geometrical approach and robust ellipsoid fitting algorithm. J. Food Eng. 358 (July), 111656 https://doi.org/10.1016/j.jfoodeng.2023.111656.

Xiong, Zilan, et al., 2021. Yi Zhong Ji Yu Wu Lian Wang De Zhi Zhu Shu Ju Jing Zhun Chuan Shu Fang Fa Ji Xi Tong (in Chinese).

Xu, Jiajun, Lu, Yuzhen, Olaniyi, Ebenezer, Harvey, Lorin, 2024. Online volume measurement of sweetpotatoes by A LiDAR-based machine vision system. J. Food Eng. 361 (August 2023), 111725 https://doi.org/10.1016/j.jfoodeng.2023.111725.

Zhang, Yifan, et al., 2021. A comprehensive review of one-stage networks for object detection. In: Proceedings of 2021 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC, pp. 2–7, 2021.

Zhu, Jun-Yan, Park, Taesung, Isola, Phillip, Efros, Alexei A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2242–51. http://ieeexplore.ieee.org/document/8237506/.