International Journal of Precision Engineering and Manufacturing (2024) 25:409–428
https://doi.org/10.1007/s12541-023-00911-7

Online ISSN 2005-4602
Print ISSN 2234-7593

REVIEW

The Vision-Based Target Recognition, Localization, and Control for Harvesting Robots: A Review

Jingfan Liu[1] · Zhaobing Liu[1]

Received: 20 June 2023 / Revised: 22 September 2023 / Accepted: 25 September 2023 / Published online: 14 October 2023

Abstract
In recent years, the elderly population has increased, leading to a labor shortage and the increasing cost of training experienced labor. Owing to the continuous optimization of machine vision, multi-sensor technologies, control methods, and end-effector structures, harvesting robots have experienced rapid development. However, most harvesting robots still require intelligent solutions, and the lack of integration with artificial intelligence limits them to small-scale applications without mass production. This paper reviews key technologies for vision-based sensing and control of harvesting robots, focusing on potential applications of vision for target recognition and localization in complex agricultural environments, describing improved solutions for different target detection and localization algorithms, and comparing their detection results. The challenges and future trends of applying these key vision sensing and control techniques in harvesting robots are also described and discussed in this review.

Keywords Harvesting robot · Target recognition · Target localization · Deep learning · Vision-based control

1 Introduction

Fruits are essential to a healthy lifestyle due to their high concentration of vitamins [1] and essential fiber [2]. The nutrients in fruits have various health benefits such as disease prevention, and this is why people buy them daily to promote well-being [3, 4]. Therefore, the growing demand for fruits necessitates continuous production and supply. Currently, most fruits are still harvested manually because this task requires the knowledge and experience of seasoned orchard workers. The manual harvesting process often results in mistakes, omissions, and damage to fruits. Consequently, this method of harvesting has resulted in increased costs throughout the agriculture industry [5] and has also worsened yield depression [6]. In the agricultural sector, modern technology, such as the use of harvesting robots, can assist farmers in overcoming these challenges and increasing their level of productivity. In recent years, computer vision technology has been implemented in harvesting fruits to detect and locate produce more efficiently [7–10].

The computer vision system initially captures raw image data through sensors or cameras, in which feature extraction [11], machine learning [12–14], and deep learning techniques [15–17] are employed to segment and detect fruit images. Once the location of the target is detected, it becomes the input for the control system. The manipulator then moves according to the planned trajectory, ultimately grabbing the target with the end-effector. The control system receives the input signal again to guide the manipulator to the next target.

Researchers have developed various types of harvesting robots that offer a novel approach to intelligently harvesting fruits. However, the unpredictability of their performance, low efficiency, and high costs currently prevent the large-scale adoption and replacement of skilled orchard workers. Robotic grasping has been proven inaccurate and inefficient due to several reasons: fruit obscured by branches and leaves in unstructured orchard environments [18], environmental factors such as illumination changes, wind, and rain that interfere with robot functionality and contact with leaves [19], and inadequate color differentiation between the orchard background and fruit [20, 21].

✉ Zhaobing Liu
zhaobingliu@whut.edu.cn

[1] Hubei Digital Manufacturing Key Laboratory, School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China

To overcome the objectively imposed disturbances, researchers have increased the accuracy of fruit recognition and localization, as well as the performance of vision-based control techniques, to improve the overall performance of the robot harvesting [22]. In this paper, the techniques proposed by researchers are broadly classified into the following three categories:

(1) Traditional image processing techniques are based on low-level features such as color, texture, and shape [23];
(2) Classification algorithms that are based on machine learning, such as the Bayesian classifier algorithm [24], Support Vector Machine (SVM) algorithm [25], K-Nearest Neighbors (KNN) clustering algorithms [26], and so on;
(3) Object detection algorithms that are based on deep learning, such as Convolutional Neural Networks (CNN) [27–29], Faster Regions with Convolutional Neural Networks (Faster R-CNN) [30–32], You Only Look Once (YOLO) network [33–36], Fully Convolutional Network (FCN) [37–39], and Single Shot MultiBox Detector (SSD) Network [40–42], etc..

It is worth noting that deep learning has become the favored technique for contemporary agricultural researchers to identify and detect fruit, thereby replacing conventional machine learning algorithms. Conventional machine learning algorithms require a manual feature extractor to extract underlying features like color, texture, and shape from image data, which is highly complex and time-consuming [43]. Data and algorithms are interdependent in computer vision, and deep neural networks require large amounts of high-quality data. Insufficient high-quality data makes the generation of an ideal model difficult, even with advanced algorithmic training. Various sensors are used for image data acquisition, such as black-and-white, RGB, hyperspectral, multispectral, and thermal cameras. Most researchers prefer using RGB cameras for image acquisition, which solely provides 2D location information without real-world location mapping [44–48]. To obtain depth information for fruit localization, researchers measure depth indirectly through binocular stereo-vision methods or physical means [44]. Previous results have reported the development of vision-based control technology with application to robotic harvesting, however, the low successful rate of fruit recognition, inefficiency localization, and inaccurate control limit the performance of harvesting robots. Therefore, a review of vision-based sensing and control technology is necessary to promote further developments for harvesting robots.

This paper presents a methodical review of recent vision-based research on harvesting robots, intending to propose solutions for target recognition and hand–eye coordination control. To provide a comprehensive overview of the topic, the subsequent sections of this paper are organized as follows: Sect. 2 introduces the key components of harvesting robots, and Sect. 3 discusses fruit detection and identification techniques in detail. Furthermore, Sect. 4 presents localization methods for fruits and their associated sensors, while Sect. 5 focuses on vision control techniques for the harvesting robot. The challenges and future trends of harvesting robots are highlighted in Sect. 6. Finally, Sect. 7 provides conclusions.

2 Key Components of Harvesting Robots

Fruit detection, positioning, and separation are three fundamental tasks that harvesting robots must execute [45]. The robotic system employs sensors to collect environmental data, which identifies and locates the target fruit. The robot control scheme then utilizes this data to maneuver the manipulator to reach the cutting point of the fruit for harvesting [46–48]. In addition, machine vision systems recognize and locate the fruit, enabling precise control of the movements. Ultimately, the manipulator and end-effector operate in tandem to divide and harvest the fruit [49]. This section describes the primary components of harvesting robots mentioned earlier. As depicted in Figs. 1 and 2, the three central technical components of harvesting robots in a laboratory or natural environment are highlighted.

2.1 Machine Vision System

Identifying the target fruit is the primary task of a harvesting robot, and taking the position information of the detected fruit as input to the robot control system is a key step in robot movement. The unstructured ambient circumstances of orchards, as well as the shadowing of fruit by tree canopies, provide massive issues for machine vision systems.

Over the past few decades, researchers have developed and deployed various machine vision-based methods for fruit detection [57–59]. Standard image identification systems are highly sensitive to light fluctuations and require costly sensors to produce high-quality images. As deep learning applications in agriculture continue to expand, researchers have focused on exploring and validating novel algorithms to tackle the previously described issues [60–62]. Studies have utilized diverse techniques, e.g., monocular cameras, binocular stereo vision cameras, RGB-D cameras, and ground laser scanners, to provide depth information and accurately locate the fruit. Refer to Sects. 3 and 3 for detailed insights into methods for fruit detection and localization.

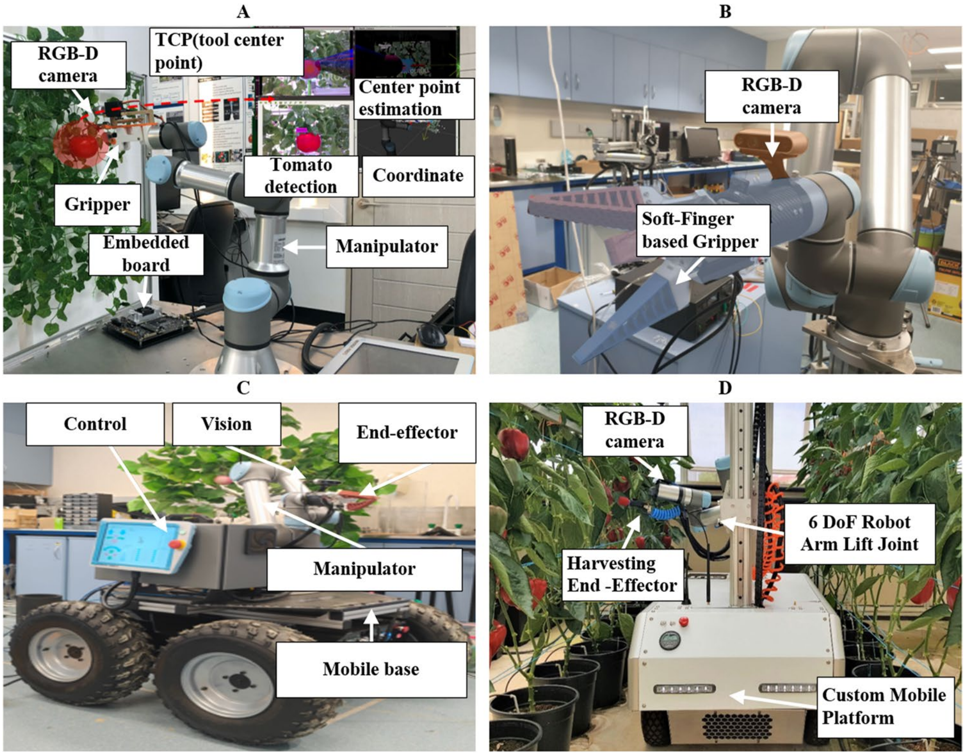


Fig. 1 Harvesting robots in the laboratory: **A** Tomato harvesting robot (adapted from [50]); **B** and **C** Apple harvesting robots (adapted from [51]); **D** Sweet-pepper harvesting robot (adapted from [52])

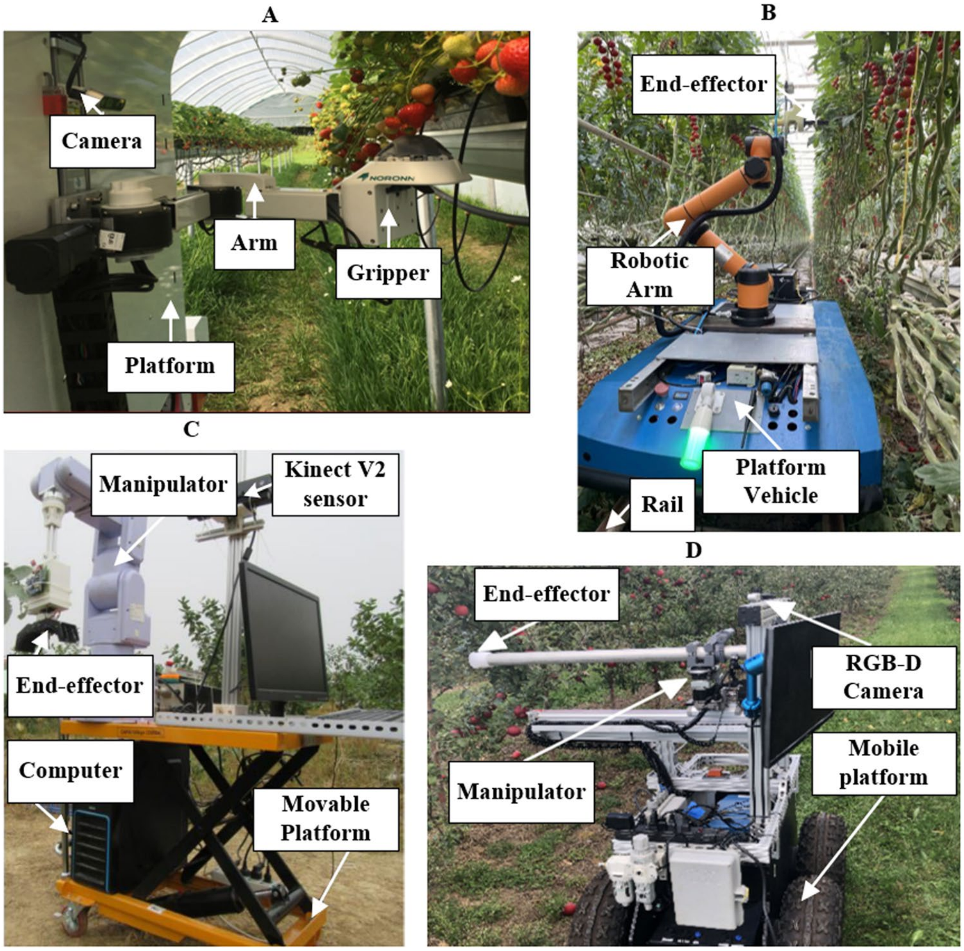


Fig. 2 Harvesting robots in the external natural environment: **A** Strawberry harvesting robot (adapted from [53]); **B** Cherry-tomato harvesting robot (adapted from [54]); **C** Guava harvesting robot (adapted from [55]); **D** Apple harvesting robot (adapted from [56])

2.2 Manipulators

Researchers have conducted extensive studies in the fields of manipulator path planning and motion planning. Maneuvering the manipulators towards the target fruit while finding the optimal cutting direction for the end-effector is a significant challenge because of the complexity and variety of the environment, including branches, leaves, and other obstructions surrounding the fruit, as well as strong winds that can arise at any time [63–65]. Manipulators have degrees of freedom and joint types (rotational or prismatic), which significantly impact kinematic flexibility, obstacle avoidance, and the space requirements needed to obtain the desired position and orientation for the end-effector during maneuvering. Recently, researchers have utilized vision-based, servo-controlled manipulators for movement planning, enabling them to harvest fruits while avoiding damaging their surroundings and navigating around obstacles [50]. According to the design requirements of the manipulator robot arm, the harvesting robot manipulator arm can be decomposed into different degrees of freedom, of which the linkage parameters and joint position parameters can determine six degrees of freedom (rotation, translation, and slewing) [66]. The trajectory of the manipulator arm is obtained by interpolating the trajectory of the joint space. By transforming the coordinate system of the six joints in the area, it can be mapped to the right-angle coordinate system of the workspace, and then using the Lagrangian method, the force that can be withstood by the six joints in the process of movement is solved, to obtain the relationship between the force exerted on the six joints of the robotic arm and the parameters of the joints.

2.3 End-Effectors

The last operation for a harvesting robotic system is fruit harvesting with the end-effector. To satisfy the requirements of actual applications, the end-effector is designed with the following aspects in mind:

(a) Reasonable gripping force. Excessive force would damage the stem and destroy the orchard condition;
(b) Harvesting efficiency;
(c) Circulation time;
(d) The sensible structure that avoids damage to the fruit and canopy structure due to the bulky mechanical parts of the end-effector [56].

For fruit harvesting, researchers have invented end-effectors with various shapes and sizes. There are two versions of automatic harvesting methods: (1) End-effectors apply mechanical force (twisting, stretching, or bending) to the fruit to separate it from the stem. (2) New cutting techniques are being sought that cut the peduncles immediately when the end-effector grips the fruit, as certain fruits have hard peduncles that make detachment difficult [67]. Soft robot end-effectors are increasingly replacing rigid end-effectors in robotic systems. They can bend to match the angle of the fruit, helping to prevent mechanical collisions with the tree canopy and trellis wires, and are more capable in that regard [68].

3 Detection Approaches for Harvesting Robot

Although harvesting robots and deep learning techniques have advanced significantly in recent years, controlling robots to detect fruits in unstructured orchards still requires considerable effort [69]. To develop classifiers, researchers gather low-level features such as textures, colors, and shapes, and then use machine learning techniques such as the SVM algorithm, K-means clustering algorithm, and AdaBoost algorithm to detect and classify fruits. Unlike traditional machine learning, deep learning allows for the creation of more abstract, high-level features or attribute categories that can improve accuracy [70]. This section mainly introduces traditional machine learning-based image processing methods and deep learning-based image identification methods. Figure 3 demonstrates the use of computer vision technology in a strawberry-picking robot to identify strawberries at different levels of ripeness.

3.1 Image Processing Techniques Based on Machine Learning

Due to the constantly changing backdrop and illumination of fruits, researchers often use extracted low-level features to segment and detect target fruits. The flow and methods

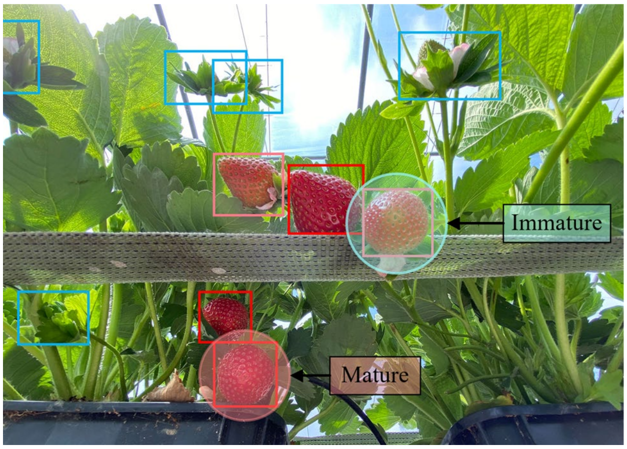


Fig. 3 The use of computer vision technology in strawberry-picking robot (adapted from [71])

of traditional image recognition technology, which are illustrated in Fig. 4, involve image preprocessing that eliminates extraneous information, recovers useful information, improves detectability, and simplifies data to enhance feature extraction, image segmentation, matching, and detection reliability. Images are acquired using various types of sensors, including black-and-white cameras [72], RGB cameras [73–75], hyperspectral cameras [76–78], multispectral cameras, and thermal cameras [79–81], multispectral cameras [79, 80] and thermal cameras [81–83]. Image data is then preprocessed using color space transformation, histogram equalization, and noise reduction techniques [43, 84, 85]. The majority of the images are captured using RGB cameras. Any additional types of sensors involved in image fusion methods will be discussed individually.

3.1.1 Detection Algorithm Based on Color Features

Zhang et al. [86] proposed a color-based technique for detecting cucumber fruits in greenhouses, achieving a 76% detection rate for mature fruits. The identification rate was hindered by the misclassification of fruit with high highlights on the surface as leaves, as well as the exclusion of partially occluded fruit due to its categorization as noise. To counteract the aforementioned issues with light and shadows, Fan et al. [87] presented a pixel block segmentation approach based on a gray-centered red–green–blue (RGB) color space, which effectively distinguishes apple-fruit pixels from other pixels, such as shadow areas. Jidong et al. [88] developed a color feature-based recognition approach to solve the issue of overlapping apples. However, the identification rate for obscured apple fruits was only 86%, highlighting the need for further improvement. Identifying unripe fruit is crucial for farmers to optimize fertilizer application during the ripening phase and forecast yield before harvesting. Zhao et al. [89] presented an algorithm for immature green-orange detection that employs color features and an absolute transformation difference. After classification and detection using the Support Vector Machine (SVM) classifier, the algorithm achieved an accuracy of over 83%. Tan et al. [90] utilized histograms of gradient orientation and color features to differentiate blueberry fruits that vary in maturity. The authors compared the accuracy of the K-Nearest Neighbor (KNN) classifier to the newly developed Template Matching with Weighted Euclidean Distance (TMWE) classifier and determined that the TMWE classifier achieved higher accuracy at a lower computational cost.

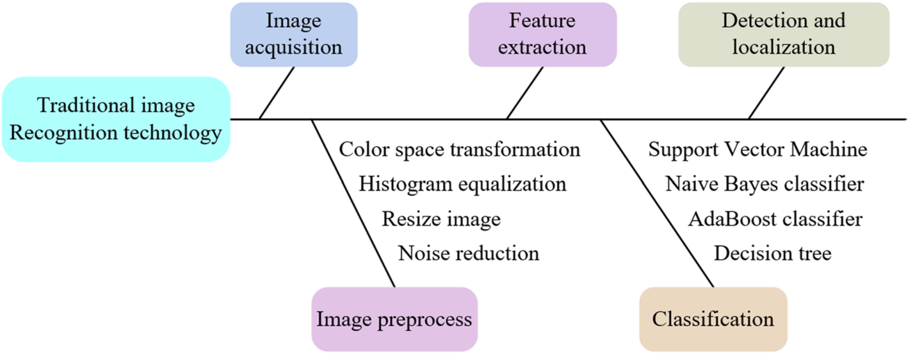
3.1.2 Detection Algorithm Based on Geometric Features

Geometric features based on shape and size can be utilized to detect apricot species using various machine-learning-based algorithms. Yang et al. [91] proposed an approach that utilized different algorithms, like decision trees, KNN, Naive Bayes, linear discriminant analysis, SVM, and artificial neural networks. The authors reported that SVM with a continuous projection algorithm led to the most accurate detection. Additionally, Lin et al. [92] introduced a novel approach for detecting apricot species by first generating a shape descriptor through contour information-based feature detection for partial shape matching. Then, the probabilistic Hough transform was used to locate candidate fruits, and lastly, the SVM classifier was used to identify all candidate fruits.

3.1.3 Detection Algorithm Based on Texture Features

In terms of texture features. Yamamoto et al. [93] proposed a machine-learning algorithm to detect various types of tomato fruits, including ripe, immature, and young fruits, by fusing multiple features. By calculating the appropriate number of clusters with X-means clustering, the algorithm detected individual fruits. However, the recall rate for young fruits was just 78%, highlighting the difficulty of distinguishing them from the stems due to size and appearance similarities. Li et al. [94] introduced a fast normalized cross-correlation (FNCC) machine vision-based algorithm to identify immature green citrus fruits by minimizing the impact of lighting fluctuations on RGB images. The algorithm combined color, shape, and texture features, with the

Fig. 4 Traditional image recognition technology



KNN classifier achieving 84.4% detection. Additionally, the authors suggested manually adjusting the camera shutter set to produce a more consistent image brightness for better fruit-to-leaf differentiation. Zhang et al. [95] researched a color and texture fusion feature-backed approach to improve apple image segmentation. The Random Forest classifier, with a 94% accuracy rate, outperformed the other eight machine learning algorithms tested for pixel classification.

3.1.4 Multi-feature Fusion Method

Regarding multiple features, Lin & Zou et al. [96] introduced a novel segmentation method that made use of an AdaBoost classifier and texture-color features that fused Leung-Malik texture and HSV color features to detect citrus by fixed-size sub-windows. Nevertheless, the LM filter bank was impacted by fluctuations in light, resulting in the over-segmentation of citrus images. Additionally, Wu et al. [97] introduced a method for detecting juicy peaches that makes use of color data and three-dimensional (3D) contour features. The study utilized a conditional Euclidean clustering approach to cluster preprocessed 3D point clouds of fruit trees. In addition, 3D contour features were used to locate and harvest the fruits. Unfortunately, when detecting unripe green fruits, the accuracy of the method was relatively low. To address this issue, Wu et al. [98] proposed a fruitful point cloud segmentation approach that blends 3D geometric features with color features. This new method demonstrates superior performance compared to the traditional color segmentation method, with an accuracy of 80.09%. Although the method is effective in detecting fruits with roughly round or spherical surfaces, the authors caution that it may not be reliable for image segmentation of fruits with irregular surfaces.

It is noted that the above image processing algorithms can be summarized in Table 1. It is further concluded that color can be used as the main extracted feature when the color of the fruit is distinguishable or it is more differentiated from the background color, such as apricots, peaches, and citrus crops with more obvious colors. However, color features rely too much on the ideal situation of light, so color extraction is usually performed under artificial conditions. When the color of the fruit is similar to its background, the shape feature can be used as the main extracted feature, e.g., green-colored fruits are similar to the color of the branches and leaves, and their shapes can be detected to improve recognition accuracy. When branches or clusters heavily occlude the fruits, texture features can be used to recognize the target fruits more quickly and accurately. By extracting multiple features, the accuracy of target recognition and adaptability to complex real-world environments can be significantly improved, and the constraints under non-artificial conditions can be reduced.

3.2 Image Recognition Technology Based on Deep Learning

Deep learning constitutes a subset of machine learning techniques [99, 100]. In traditional machine learning algorithms, the ability to learn is typically constrained, and larger amounts of data do not necessarily result in continuous improvement in the information learned. Conversely, deep learning systems, like the machine equivalent of "more experience," are capable of enhancing performance by accessing vast amounts of data. To overcome the challenge of numerous parameters and lengthy optimization times, the advent of GPU parallel computing technology has triggered a global rise in deep learning research. Additionally, comprehensive and rigorous investigations on the application of deep learning to agricultural robots have been conducted.

3.2.1 Two-stage Object Detection Algorithm

Faster R-CNN is a typical two-stage object detection model, and its structural diagram is shown in Fig. 5. The RPN (Region Proposal Network) is a crucial innovation that connects the region generation and convolutional networks through an anchor mechanism. It achieved an increased

Table 1 Machine learning based image processing algorithms

Detection algorithm	Advantage	Disadvantage	Detection rate (%)	Detection target	References
Based on color features	Easy to operate, distinguishing fruits and backgrounds	More seriously affected by the similarity of illumination intensity and background color, etc	80–85	Citrus, apples, peach	[87–90]
Based on geometric features	Obtainoutline information of fruits without the influence of illumination	Affected by shading from branches, limbs, clusters, etc	80–87	Green apple, green citrus	[91, 92]
Based on texture features	More viable under occlusion than before color and geometric features	Influenced by factors such as light intensity and environmental obscuration	75–90	Pineapple, bitter gourd	[93–95]

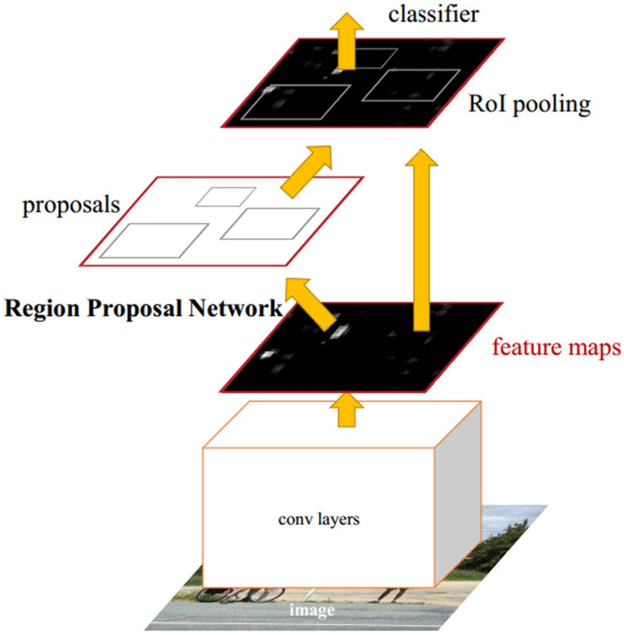


Fig. 5 Structure diagram of Faster R-CNN (adapted from [30])

detection rate of 17 frames per second [30]. With the Faster R-CNN as the foundation, Kaiming [101, 102] introduced the Mask R-CNN, an innovative instance segmentation network. The key improvement can be summarized as:

(1) To resolve the accuracy loss caused by the ROI pooling rounding method, the RoI Align method was introduced as a replacement for the original ROI pooling method.

(2) A mask branch was incorporated into the image segmentation model to determine the class of each pixel.

Furthermore, new and improved algorithms based on the two-stage detection algorithm have been introduced to meet the requirements of various fruit detection tasks. Jia et al. [103] proposed a Mask Region Convolutional Neural Network (Mask R-CNN) based visual detector model for harvesting robots. The authors tested the method with a random test set of 120 images and achieved 97.31% accuracy and 95.70% recall. Parvathi et al. [104] have proposed an improved faster region-based convolutional neural network (Faster R-CNN) model for the detection of two important ripening stages of coconut in a complex context. The Faster R-CNN algorithm based on the ResNet-50 network was used to improve the detection scores of nuts at the two major ripening stages. Table 2 provides an overview of the improved two-stage target detection algorithms.

3.2.2 One-stage Object Detection Algorithm

The two-stage object detection algorithm creates region proposals in the first stage. In the second stage, the contents of the region of interest are classified and regressed, but this results in the omission of spatial information for local objects within the entire image. To solve this problem, a one-stage object detection algorithm is proposed that omits the region proposal creation stage and can directly detect objects. One of the most representative and popular single-stage target detection algorithms is the YOLO series [113–116], whose structure is shown in Fig. 6. The YOLO series has a faster detection speed than the R-CNN series,

Table 2 Improved two-stage object detection algorithm

Target	Description	Precision (%)	References
Apple	ResNet and a Densely connected convolutional network were applied to form the backbone network. In this context, the dense block preserves certain low-dimensional features	97.31	[103, 105]
Coconut	ResNet-50 was one feature extraction network that incorporated the idea of the Residual Network (ResNet)	89.4	[104, 106]
Cucumber	Resnet-101 was chosen as the backbone of Mask R-CNN, and the logical green (LG) operator was introduced as a filter for non-green backgrounds	90.68	[107]
Grape cluster	The efficient Channel Attention (ECA) mechanism was incorporated into the backbone, and Dense Upsampling Convolution (DUC) was deployed to compensate for feature information in FPN	N/A	[108]
Kiwifruit	Based on the Visual Geometry Group (VGG-16) network, an image fusion and feature fusion model were developed	90.07	[109, 110]
Mango Orange	Loss functions were applied to convolution and pooling layers, which reduced the weights of various high-dimensional parameters	88.94 90.73	[8]
Sweet pepper	Early or late fusion methods were implemented to merge multimodal information, including RGB and infrared information	83.8	[7]
Strawberry	Feature Pyramid Network (FPN) was introduced into the backbone network to strengthen the extraction of fruit features at multiple scales	95.78	[111]
Tomato	Edge contour detection was combined with deep learning	80	[112]

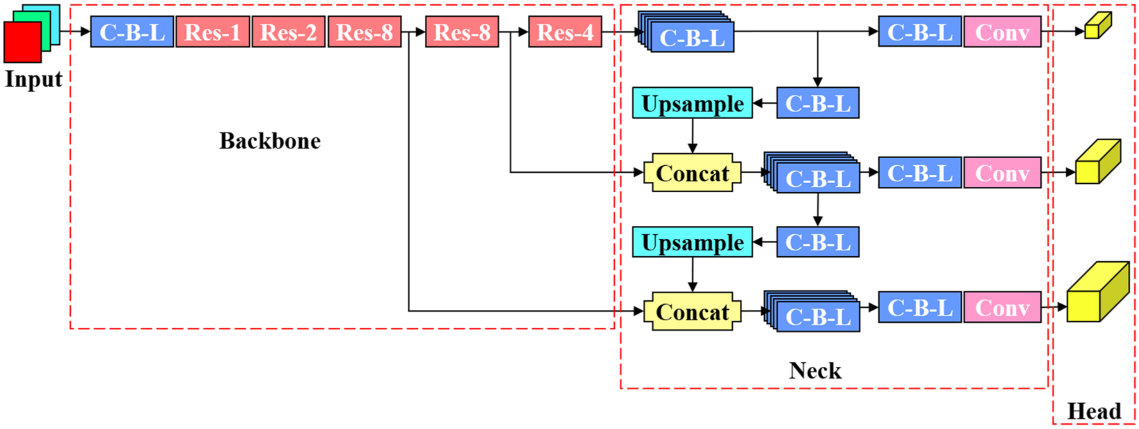


Fig. 6 Structure diagram of YOLO-v3

but the detection accuracy of the YOLO series is usually inferior to that of the R-CNN series.

To balance detection accuracy and speed for optimal gains, another one-stage object detection algorithm called SSD was presented by Liu et al. [42]. During the last five years, several improved algorithms based on the YOLO or SSD framework have been developed. Tian et al. [117] have introduced an improved YOLOv3 model for detecting apples at different growth stages in orchards with fluctuating light, complex backgrounds, overlapping apples, and complicated branches and foliage. The proposed YOLOv3 dense model outperforms the original YOLOv3 model and the Faster R-CNN based on the VGG16 network, with an average detection time of 0.304 s/frame, which enables real-time detection of apples in orchards. To automatically identify graspable and non-graspable apples in apple tree images, Yan et al. [118] proposed a lightweight target detection method for an apple-picking robot based on improved YOLOv5s. The experimental results show that the improved network model can effectively identify graspable apples that are not occluded by leaves or only occluded by leaves, as well as non-graspable apples that are occluded by branches or occluded by other fruits. Specifically, the recognition recall, precision, mAP and F1 were 91.48%, 83.83%, 86.75% and 87.49%, respectively. The average recognition time was 0.015s per image. Overall, the improved one-stage object detection algorithm used in fruit harvesting robots is outlined in Table 3.

4 Localization Methods for Harvesting Robot

Harvesting robots require 3D spatial position information from the detected fruit to guide the end effectors through the harvesting procedure. However, the camera obtains only the 2D image space position of the target, thus needing to establish a mapping relationship between the target position in the 2D image space and the 3D space position. Researchers have proposed successful solutions to address this issue, which is introduced in this section. The paper categorizes the methods for localizing fruit based on camera data into 2D and 3D categories. The detailed comparison of 2D and 3D cameras is shown in Table 4.

4.1 Localization Method Based on Two-dimensional Images

2D cameras that contain charge-coupled device (CCD) sensors or complementary metal oxide semiconductor (CMOS) sensors are prevalent for fruit localization and the trajectory tracking of harvesting robots [146]. Mehta et al. [134] acquired 3D positioning information on citrus fruits using a monocular camera based on perspective transformation. The authors demonstrated that this method was less temporally complex than a stereo vision technique's depth estimation method after comparing test results. Xiong et al. [135] used a CCD camera with artificial lighting to detect green grapes and locate harvesting points, preventing the missing and inadvertent collection of nascent grapes during the night. They found that the highest accuracy in harvesting point detection was 92.5% at a depth of 500 mm. However, they also discovered that increased shot distance reduced light density, causing errors in point computation due to poor image quality.

Accurate fruit localization is crucial for effective robotic harvesting. Mehta et al. [136] developed a nonlinear estimator based on particle filters to predict fruit locations captured from multiple CMOS cameras. However, the approach has limited effectiveness in the case of an obstructed view. Unpruned buds can hinder accurate localization by producing vegetation that conceals new buds and affects the

Table 3 Improved one-stage object detection algorithm

Target	Description	F1 (%)	References
Apple	Utilized DenseNet as a substitute for the original transport layer to improve network performance	81.7	[117]
	Designed the BottleneckCSP-2 module based on the improvement of the BottleneckCSP module, and embedded the SE module into the backbone	87.49	[118, 119]
	A bi-directional feature integrated Pyramid Network Small (BiFPN-S) network was added to the backbone, and the Activate Or Not (ACON)-C activation function, which replaced the SiLU function, was implemented	92.8	[120, 121]
Citrus	SimAM attention mechanism module is added before each feature detection layer. Combined with the Canopy algorithm and k-means + + algorithm to obtain more matching anchor boxes	91.95	[122, 123]
	An image fusion algorithm based on an improved Laplace pyramid was introduced, and ResNet units were embedded in the original network	93.56	[124]
Cherry tomato	The dual path network replaced the original feature extraction network and increased the 104 104 resolution of the feature layers in the FPN. An improved K-means + + clustering algorithm was utilized to calculate the scale of the anchor box	94.18	[125]
Grape	Squeeze-and-Excitation Networks (SE) attention mechanism was joined to the network, and non-maximum suppression (NMS) was replaced with soft NMS	90.47	[119, 126, 127]
Green pepper	GhostConv was applied to change the Conv layer of Cross Stage Partial (CSP). A single-layer structure, BiFPN, was chosen instead of Path Aggregation Network (PANet)	78.9	[128, 129]
Kiwifruit	Adopted MobileNetV2 to displace vggnet-16 as a feature extractor. And quantized the model to obtain faster deduction with smaller model size	N/A	[130, 131]
Longan	The core component of the MobileNet model was switched out for a feature extraction network, which decreased the amount of time airborne computers needed to calculate and detect things	N/A	[132]
Tomato	A parallel sub-network, RGB-Network, was augmented to the previous SSD framework, which integrated multimodal features and generated accurate feature maps	91.47	[133]

Table 4 The detailed comparison of 2D and 3D cameras

	Typical sensors	Advantages	Disadvantages	Measurement range (m)	Error (mm)	References
CCD	MV-E800C, KPC-S20-CP1	Outperforms CMOS sensors in sensitivity, resolution, and noise control and provides color, geometry, and texture information	Only provides two-dimensional information and is easily influenced by light	0.3–1.6	< 7.5	[134, 135]
CMOS	C920-Pro, Galaxy A5	Low cost, low power consumption, and high integration. Provides color, geometry, and texture features	Only provides two-dimensional information and is easily influenced by light	–	< 15	[136, 137]
Structured Light	Kinect-1414	High accuracy when measuring objects at close range	The accuracy of measuring objects at long distances is low. Affected by light and light reflection	0.2–3.5	< 10	[138]
Stereo	ZED-2, Intel-D435	High accuracy when measuring objects at long range; less affected by light	High hardware cost with low image resolution and high power consumption	0.11–10.0	< 2%	[139, 140]
Time of Flight	Kinect-v2, Mesa-SR4000	Compact, lightweight, high image resolution	Poor real-time performance and high computational costs	0.1–6.0	< 8	[141–145]

subsequent output. To address these issues, Daz et al. [137] developed a grape bud detection and localization method based on motion structure and 2D image categorization. The approach captured 2D images to construct a 3D model of the scene, achieving a localization error of 259–554 pixels.

4.2 Localization Method Based on Three-dimensional Coordinates

Conventional RGB color cameras, known as 2D cameras, only capture objects in the sensor field of view and cannot acquire component depth information. However, RGB-D depth cameras can acquire this depth information directly. The depth camera captures data and calculates the distance of each point in the picture from the camera. The x and y coordinates combine to provide the 3D spatial coordinates of each point in the image. Researchers have used depth cameras in combination with robotics to create more efficient robotic harvest systems. Depth cameras are categorized as structured light cameras, binocular cameras, and Time of Flight (ToF) cameras based on their operating principles.

4.2.1 Localization Based on Structured Light

A structured light camera consists of a laser projector and one or more structured light cameras. The projector actively emits infrared light onto the object's surface, which is then imaged with one or more structured light cameras. By calculating the location and depth information based on the triangulation principle, 3D reconstruction is achieved [147]. Laser triangulation is a fundamentally structured light system, as shown in Fig. 7.

Structured-light cameras are widely utilized in agricultural automation. Nguyen et al. [138] used an RGB-D structured light camera to acquire images and developed an algorithm based on color and shape features that detected and located red and bi-colored apples beneath an umbrella blocking direct sunlight. The positioning accuracy in all directions was less than 10 mm. Additionally, the authors suggested using additional sensors for more information on the 3D position of the fruit and the location of the stem to enhance the gripping and harvesting of individual fruits.

4.2.2 Localization Based on Binocular Stereo Vision

Binocular stereo vision involves taking two pictures of the object of interest using cameras placed at different locations and then determining the positional difference between the corresponding points in the two images to obtain 3D geometric information about the object [147]. Figure 8 shows its schematic diagram. The system for binocular stereo vision is constructed using two conventional consumer-grade RGB cameras due to the inexpensive nature of the camera hardware requirements. Wang et al. [148] presented a technique for target localization using window scaling. The approach involves collecting photos of produce and estimating the three-dimensional coordinates of the target by utilizing the triangulation principle to achieve complete target localization. In a natural environment, Liu et al. [149] implemented a binocular stereo-vision approach and an improved YOLOv3 model for pineapple identification and localization. At a range of 1.7–2.7 m, the absolute mean error was 24.414 mm, with an average relative error of 1.17%. Additionally, during robotic harvesting, the method took into account wind disturbance, mutual branch contact, and mechanical collisions. The visual localization of dynamic lychee clusters was explored by Xiong et al. [150] The harvesting point was determined by computing the oscillation angle of lychee clusters in three states of disturbance: static, slight, and large. The maximum depth error was 5.8 cm, and the minimum depth error was 0.4 cm.

In practice, biocular depth cameras have been widely adopted by researchers for developing vision systems for harvesting robots. Hou et al. [139] reported a recent technique utilizing modified YOLOv5 and binocular stereo vision for detecting and localizing ripe citrus. The average distance error between citrus fruit and the camera in non-uniform, low, and good lighting conditions was 3.98 mm. The approach was found to offer accurate and swift detection and localization of citrus fruits in intricate orchard landscapes, as concluded by the authors. Occlusion of leaves,

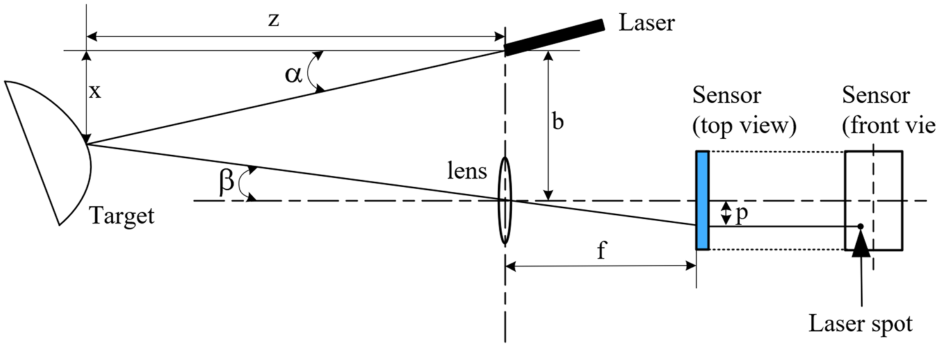


Fig. 7 Triangulation with a single laser spot (adapted from [147])

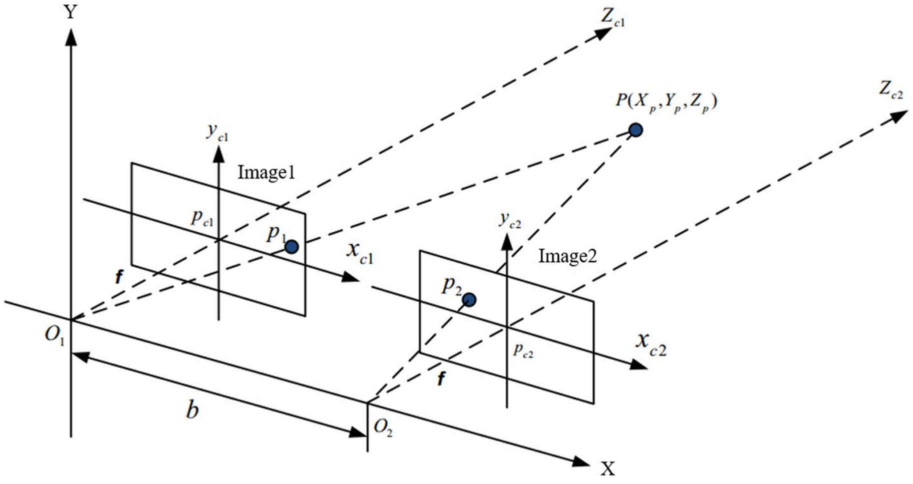


Fig. 8 Schematic diagram of binocular stereo vision (adapted from [151])

branches, and other fruits often leads to imprecise bounding boxes of detected fruits and the associated depth measurements. To manage this problem, Li et al. [140] proposed a distinctive 3D fruit localization method dependent on a truncated cone point cloud processing algorithm. According to the authors, this method decreased the median and average fruit localization errors by 59% and 43%, respectively, when compared to the traditional approach. Wang et al. [152] presented a geometry-aware detection network designed for apple harvesting. The network utilized color and geometry sensory input from RGB-D cameras and executed end-to-end instance segmentation and grasping estimates with an average precision of 0.61 cm and 4.8° in center and orientation, respectively.

4.2.3 Localization Based on the Principle of ToF

A time-of-flight camera consists of a light transmitter and a receiver. The receiver detects the light emitted by the transmitter once it reflects off the object in view, and the distance to the target object is determined by measuring the time taken for the signal to travel to and reflect off the object [147]. Wu et al. [141] developed a platform of robotic devices resembling bananas that utilize stereo vision to improve 3D localization accuracy at the truncation point. The study found a median error of 8 mm and a median absolute deviation of 2 mm for the depth coordinates. Lin et al. [55] utilized Euclidean clustering guided by fruit binary maps and RGB-D depth pictures to separate point clouds into individual fruits to enable collision-free harvesting. By determining the center location of each fruit and its relation to the mother branch, it was possible to accurately estimate the 3D poses of the fruits. The study found that the 3D posture error, calculated using the spherical fitting method, was 23.43° ± 14.18° and that the method took 0.565 s to execute per fruit. Li et al. [143] employed principal component analysis (PCA) to estimate the positioning for lychee harvesting, specifically targeting the random scattering and uneven appearance of lychee clusters. The study achieved a detection precision of 83.33% and a placement precision of 17. 29° ± 24.57°. Therefore, further improvements in accuracy were deemed necessary by the authors.

5 Vision-Based Control for Harvesting Robot

Vision servo control is a widely used robotics technique that utilizes vision sensors to gather environmental data and translate it into appropriate kinematic commands for the controller of robots. Early robot vision systems utilized open-loop vision control, employing a "look then move" approach rather than employing closed-loop control. Ongoing advancements in computer hardware and related algorithms have led to the recent and rapid development of technology in this field, as shown by recent research [153–155]. This section presents an overview of the control methods utilized in vision-based harvesting robots, as discussed previously. Table 5 details the various control methods and overall performance metrics of harvesting robots utilized in prior years.

5.1 Open-loop Visual Control

Silwal et al. [158] employed RGB-D cameras to develop a robot for apple harvesting with open-loop vision control. The study found a successful harvesting rate of 0.846 and attributed the partial failure to progressive position errors in open-loop vision control systems and difficulty catching apples on long, thin, flexible branches. Ling et al. [156] developed a dual-arm cooperative technique utilizing binocular vision sensors to improve the efficiency of tomato harvesting robots, which involved tomato detection, target localization, trajectory planning, and real-time control of dual-arm motions. The study achieved a success rate of up

Table 5 Control mode and performance of harvesting robot

Robot	Vision-based control	Success rate (%)	Speed (s)	References
Apple harvesting robot	Open loop visual control	72	14.6	[156]
	Image-based visual servo	77	15	[157]
	Open loop visual control	84.67	7.6	[158]
	Image-based visual servo	91.2	13.8	[159]
Cherry-tomato harvesting robot	Image-based visual servo	83	8	[160]
Citrus harvesting robot	Image-based visual servo	NR	< 8	[161, 162]
Tomato harvesting robot	Open loop visual control	87.5	< 30	[156]
Sweet-pepper harvesting robot	Image-based visual servo	NR	45	[163]

to 87.5% using suction cup grabbing and wide-range cutting during robotic harvesting. Yu et al. [156] introduced a humanoid robot designed for efficient and flexible apple harvesting, utilizing the scale-invariant feature transformation (SIFT) feature point detection and matching mechanism to identify the pixel coordinates of the optimal apple contour and target apple center. The authors reported several practical issues that adversely impact the performance of robot:

(1) All the electric motors and components of the robot operated on lithium-ion batteries that were inadequate in meeting the necessary driving force.
(2) The research assumed uniform apple size, yet apples have varying sizes, requiring the claws to be more adaptable or the addition of a tactile sensor for better performance.
(3) Although the color segmentation-based identification accuracy of binocular camera systems is unsatisfactory, it can be improved by using RGB-D cameras and advanced identification algorithms.

5.2 Visual Servo Control

Errors in the vision input device and system, localization errors in vision-based recognition, coordinate transformation errors, and other factors affect the operational efficiency of harvesting robots in an open-loop system. Cumulative mistakes fail to accurately pick certain fruits. To improve the accuracy of the robot, the visual feedback loop identifies deviations between the actual and intended positions of the manipulator [164]. Conventionally, visual servo control is either a position-based visual servo (PBVS) or an image-based visual servo (IBVS) [154, 155] depending on whether the feedback signal is a 3D spatial coordinate value or an image feature value.

5.2.1 Position-based Visual Servo (PBVS)

The position-based vision servo (PBVS) system determines the intended poses via image analysis and the geometric model of the target. The deviation between the current and goal poses informs the trajectory planning [165, 166]. PBVS control technology is built on the basis of precise measurement of the spatial coordinates of the target fruit by the visual sensor. First of all, the visual sensor must obtain accurate spatial position information of the target fruit, and then through the establishment of an accurate hand-eye coordinate conversion model, the position information under the visual sensor coordinate system will be converted to the spatial coordinates under the robot coordinate system. Finally, the positional relationship between the target fruit and the robot end-effector can be used to carry out the motion planning, which in turn can control the movement of the end-effector of the picking robot to the target fruit position. The schematic structure of the system is illustrated in Fig. 9.

5.2.2 Image-based Visual Servo (IBVS)

In image-based visual servoing, the control quantity is computed directly from the error signal in the image to drive

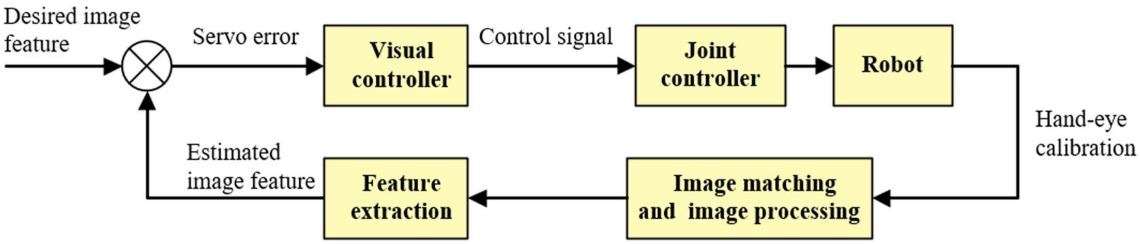


Fig. 9 Position-based visual servo

the robot to move toward the target fruit and complete the picking task. The critical problem of image-based visual servo control is the need to estimate the image Jacobi matrix, which is the bridge to construct the transformation between the image coordinate system and the robot coordinate system [165, 166]. This image-based visual servo control is relatively insensitive to robot kinematic calibration and camera model errors compared to position-based visual servo control, and does not require estimation of the position of the target fruit in the robot coordinate system, thus reducing computational latency. Therefore, it has become one of the most preferred solutions nowadays. The schematic structure of the system is depicted in Fig. 10.

In practice, image-based vision servoing is considered less computationally challenging compared to position-based servoing, making it the preferred control mechanism. Mehta et al. [161, 162] proposed a cooperative vision servo controller that addressed external disruptions, such as mechanical contact between the robot and trees, by incorporating a feedback term that compensated for positioning flaws, allowing the end-effector to micro-adjust the position. However, when the robot interacted with dense crops, it potentially missed the target, leading to harvesting failure. Barth et al. [163] designed a servo control framework to address agricultural settings with dense plant cover. The servo control framework achieved motion control of a sweet pepper harvesting robot with visual information, successfully harvesting sweet peppers under laboratory conditions. Chen et al. [159] developed a vision-based servo control for a harvesting robot using an upgraded fuzzy neural network sliding mode algorithm. The enhanced algorithm significantly increased not only the design efficiency but also the success rate of the harvest. However, the procedure was only tested in a laboratory setting, and the authors acknowledged potential obstacles when harvesting in natural settings.

6 Challenges and Future Trends

The potential of harvesting robots to revolutionize smart agriculture is immeasurable. The advancements in machine vision and artificial intelligence technologies have significantly accelerated the transition of harvesting robots from laboratory settings to practical orchards. However, fruit-harvesting robots encounter various challenges in their current implementation. These include issues associated with energy consumption and unstructured orchard environments. Fruit blockage by branches and leaves, uncertainty caused by the similarity of the background color to the fruit's body color, direct contact with fruit by bulky and inflexible robots, and the need to consider the degree of fruit ripeness and defects during harvesting are just a few examples of such challenges. Researchers must explore and enhance fruit detection, localization, and control techniques to address these problems.

6.1 Building a Structured Environment Suitable for Harvesting Robots

The haphazard growth of fruit leaves in natural environments poses a significant challenge to harvesting robots. It is often the most difficult issue for fruit-harvesting robots to detect and tackle. In recent years, contemporary garden management techniques have been employed to create structured environments that are suitable for harvesting robots. For instance, apple trees have been pruned to form a flat crown, leaves and branches are cleaned manually or mechanically, chamber agricultural systems are used to seed fruits [167–169].

6.2 Designing the End Effector Suitable for Fruit Detachment

Robots used for picking fruits are often in motion during operation, and the movement caused by wind and picking can cause the fruit to swing back and forth, resulting in damage to the fruit skin and affecting its quality. Harvesting usually involves grabbing the fruit and pulling it off the vine, which may cause mechanical damage. Therefore, designing high-precision end effectors to grab fruits is a direction for future improvement. To design an end-effector appropriate for picking robot, the following factors must be taken into account: its ability to adjust to various shapes and sizes of produce, its lightweight and flexibility to enable swift robot

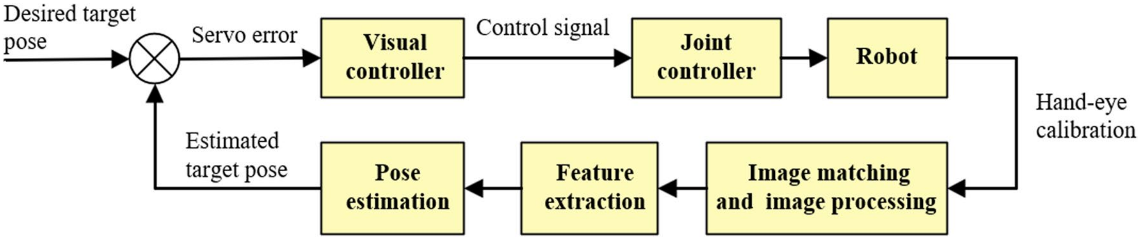


Fig. 10 Image-based visual servo

motion, and its high-precision and stability to ensure gentle and accurate picking. Moreover, the end-effector must be easy to maintain given the prolonged timeframe of operation. Therefore, it is advisable to use durable materials and a simplistic design [168, 170, 171].

6.3 Developing a More Accurate Fruit Detection and Localization Algorithm

Despite the emergence of several high-performing fruit recognition algorithms, image processing algorithms are continually being improved upon. In recent years, new vision sensors, such as light field cameras and chlorophyll fluorescence cameras, have garnered increased attention. The use of these advanced sensors to acquire higher-detail visual data would undoubtedly enhance the recognition of complex environments. The accuracy and efficiency of vision control are other areas that require improvement. Although pose-based visual servoing (PBVS) control has been applied in a variety of applications, most agricultural producers continue to employ image-based visual servoing (IBVS) control, given the economic costs and the current state of vision sensor technology. However, we should devote future efforts to implementing PBVS control [172–174].

6.4 Training a Lightweight Model for Fruit Target Detection

To enhance the recognition performance of the vision system, it is common to improve recognition algorithms. However, these methods result in a more complex algorithm and longer computing time, despite the improvement in recognition accuracy. Meeting actual production requirements becomes challenging due to the real-time nature of vision systems employed by picking robots. Consequently, priority should be given to develop lightweight target detection models that support real-time fruit detection on edge devices and enhance the performance of visual recognition systems in embedded devices [175].

6.5 Other Feasible Directions

Using multiple robotic arms can enable the efficient grasping of multiple fruits simultaneously. A collaborative effort can reduce the risk of errors and failures while improving the accuracy of grasping [144, 176]. In specific scenarios, a single robotic arm may find it challenging to accomplish the task at hand, which makes having multiple robotic arms collaborate ideal. This gives the robotic arm system more flexibility and adaptability in its application range. In the field of agriculture, visual recognition and detection technology can be incorporated into intelligent agricultural systems to help farmers achieve automated management and production. For instance, the growth status, fruit maturity, and yield of fruit trees can be monitored in real-time in the orchard through corresponding sensors. Diseases and pests can also be detected at early stages by visual recognition and detection technology, leading to reduced fruit loss and pesticide use [177]. Additionally, image analysis algorithms can automatically grade fruits based on their size and color, resulting in increased efficiency and yield quality [178, 179].

7 Conclusions

This paper presents a comprehensive review of the recent progress in fruit-harvesting robots developed by researchers in the past five years. The study discusses the advancements in target recognition and detection techniques and the methods for achieving target localization in fruit-harvesting robots. The paper compares the vision-based control techniques for harvesting robots, and after conducting a thorough survey, visual servo control is identified as the most widely used control method. The primary contribution of this paper is to provide a comprehensive and in-depth analysis of the core technologies utilized in fruit-harvesting robots. Additionally, the paper highlights significant advancements achieved through multi-sensor fusion technology, deep learning-based target detection algorithms, novel end-effectors, and vision servo-based closed-loop control, which have the potential to further enhance the intelligence, accuracy, flexibility, and efficiency of fruit-harvesting robots.

Acknowledgements This work is supported by the Open Project Program of Shandong Marine Aerospace Equipment Technological Innovation Center, Ludong University (Grant No. MAETIC2021-02).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Van Duyn, Ma. S., & Pivonka, E. (2000). Overview of the health benefits of fruit and vegetable consumption for the dietetics professional: selected literature. *Journal of the American Dietetic Association, 100*(12), 1511–1521.
2. Dreher, M. L. (2018). Whole fruits and fruit fiber emerging health effects. *Nutrients, 10*(12), 1833.
3. Siriamornpun, S., Weerapreeyakul, N., & Barusrux, S. (2015). Bioactive compounds and health implications are better for green jujube fruit than for ripe fruit. *Journal of Functional Foods, 12*, 246–255.
4. Osborne, J. D., Da Silva, M., Frace, A. M., Sammons, S. A., Olsen-Rasmussen, M., Upton, C., Buller, R. M., Chen, N., Feng, Z., Roper, R. L., & Liu, J. (2013). Fruit quality and bioactive compounds relevant to human health of sweet cherry

(*Prunus avium* L.) cultivars grown in Italy. *Food Chemistry., 140*(4), 630–638.

5. Zhang, Z., Heinemann, P. H., Liu, J., Baugher, T. A., & Schupp, J. R. (2016). The development of mechanical apple harvesting technology: A review. *Transactions of the ASABE, 59*(5), 1165–1180.
6. Bargoti, S., & Underwood, J. P. (2017). Image segmentation for fruit detection and yield estimation in apple orchards. *Journal of Field Robotics, 34*(6), 1039–1060.
7. Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. *Sensors, 16*(8), 1222.
8. Wan, S., & Goudos, S. (2020). Faster R-CNN for multi-class fruit detection using a robotic vision system. *Computer Networks, 168*, 107036.
9. Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J. R., Morros, J. R., Ruiz-Hidalgo, J., Vilaplana, V., & Gregorio, E. (2020). Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. *Computers and Electronics in Agriculture, 169*, 105165.
10. Rahnemoonfar, M., & Sheppard, C. (2017). Deep count: Fruit counting based on deep simulated learning. *Sensors, 17*(4), 905.
11. Liu, X., Zhao, D., Jia, W., Ji, W., & Sun, Y. (2019). A detection method for apple fruits based on color and shape features. *IEEE Access, 7*, 67923–67933.
12. Zhuang, J., Luo, S., Hou, C., Tang, Y., He, Y., & Xue, X. Y. (2018). Detection of orchard citrus fruits using a monocular machine vision-based method for automatic fruit picking applications. *Computers and Electronics in Agriculture, 152*, 64–73.
13. Lu, J., Lee, W. S., Gan, H., & Hu, X. (2018). Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis. *Biosystems Engineering, 171*, 78–90.
14. Tao, Y., & Zhou, J. (2017). Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. *Computers and Electronics in Agriculture, 142*, 388–396.
15. Wu, F., Duan, J., Chen, S., Ye, Y., Ai, P., & Yang, Z. (2021). Multi-target recognition of bananas and automatic positioning for the inflorescence axis cutting point. *Frontiers in Plant Science, 12*, 705021.
16. Moreira, G., Magalhães, S. A., Pinho, T., & Cunha, M. (2022). Benchmark of deep learning and a proposed HSV colour space models for the detection and classification of greenhouse tomato. *Agronomy, 12*(2), 356.
17. Zhang, W., Chen, K., Wang, J., Shi, Y., & Guo, W. (2021). Easy domain adaptation method for filling the species gap in deep learning-based fruit detection. *Horticulture Research, 8*, 119.
18. Mao, S., Li, Y., Ma, Y., Zhang, B., & Wang, K. (2020). Automatic cucumber recognition algorithm for harvesting robots in the natural environment using deep learning and multi-feature fusion. *Computers and Electronics in Agriculture, 170*, 105254.
19. Williams, H. A., Jones, M. H., & Nejati, M. (2019). Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. *Biosystems Engineering, 181*, 140–156.
20. Bac, C. W., Hemming, J., Van Tuijl, B., Barth, R., Wais, E., & Van Henten, E. J. (2017). Performance evaluation of a harvesting robot for sweet pepper. *Journal of Field Robotics, 34*(6), 1123–1139.
21. Fountas, S., Mylonas, N., Malounas, I., Rodias, E., Hellmann Santos, C., & Pekkeriet, E. (2020). Agricultural robotics for field operations. *Sensors, 20*(9), 2672.
22. Li, P., Lee, S.-H., & Hsu, H.-Y. (2011). Review on fruit harvesting method for potential use of automatic fruit harvesting systems. *Procedia Engineering, 23*, 351–366.
23. Zhao, Y., Gong, L., Huang, Y., Liu, C., et al. (2016). A review of key techniques of vision-based control for harvesting robot. *Computers and Electronics in Agriculture, 127*, 311–323.
24. Amatya, S., Karkee, M., Gongal, A., Zhang, Q., & Whiting, M. D. (2016). Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. *Biosystems Engineering, 146*, 3–15.
25. Zhang, C., Zhang, K., Ge, L., Zou, K., Wang, S., Zhang, J., & Li, W. (2021). A method for organs classification and fruit counting on pomegranate trees based on multi-features fusion and support vector machine by 3D point cloud. *Scientia Horticulturae, 278*, 109791.
26. Ghazal, S., Qureshi, W. S., Khan, U. S., Iqbal, J., Rashid, N., & Tiwana, M. I. (2021). Analysis of visual features and classifiers for Fruit classification problem. *Computers and Electronics in Agriculture, 187*, 106267.
27. Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y. D. (2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. *Scientia Horticulturae, 263*, 109133.
28. Momeny, M., Jahanbakhshi, A., & Jafarnezhad, K. (2020). Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. *Postharvest Biology and Technology, 166*, 111204.
29. Zhang, Y. D., Dong, Z., Chen, X., Jia, W., Du, S., Muhammad, K., & Wang, S. H. (2019). Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimedia Tools and Applications, 78*, 3613–3632.
30. Ren, S., He, K., Girshick, R., & Sun, J. (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems, 28*.
31. Girshick, R. (2015). *Fast r-cnn; proceedings of the Proceedings of the IEEE international conference on computer vision*.
32. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition*.
33. Junos, M. H., Mohd, K. A., Thannirmalai, S., & Dahari, M. (2022). Automatic detection of oil palm fruits from UAV images using an improved YOLO model. *The Visual Computer, 38*(7), 2341–2355.
34. Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering, 210*, 271–281.
35. Lawal, M. O. (2021). Tomato detection based on modified YOLOv3 framework. *Scientific Reports, 11*(1), 1–11.
36. Gai, R., Chen, N., & Yuan, H. (2021). A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Computing and Applications, 2021*, 1–12.
37. Barreto, A., Lottes, P., Yamati, F. R. I., Baumgarten, S., Wolf, N. A., Stachniss, C., & Paulus, S. (2021). Automatic UAV-based counting of seedlings in sugar-beet field and extension to maize and strawberry. *Computers and Electronics in Agriculture, 191*, 106493.
38. Marset, W. V., Pérez, D. S., Díaz, C. A., & Bromberg, F. (2021). Towards practical 2D grapevine bud detection with fully convolutional networks. *Computers and Electronics in Agriculture, 182*, 105947.
39. Peng, Y., Wang, A., Liu, J., & Faheem, M. (2021). A comparative study of semantic segmentation models for identification of grape with different varieties. *Agriculture, 11*(10), 997.
40. Vasconez, J. P., Delpiano, J., Vougioukas, S., & Cheein, F. A. (2020). Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. *Computers and Electronics in Agriculture, 173*, 105348.

41. Magalhães, S. A., Castro, L., Moreira, G., Dos Santos, F. N., Cunha, M., Dias, J., & Moreira, A. P. (2021). Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse. *Sensors, 21*(10), 3569.
42. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Proceedings of the European conference on computer vision*. Springer.
43. Maheswari, P., Raja, P., Apolo-Apolo, O. E., & Pérez-Ruiz, M. (2021). Intelligent fruit yield estimation for orchards using deep learning based semantic segmentation techniques—a review. *Frontiers in Plant Science, 12*, 684328.
44. Fu, L., Gao, F., Wu, J., Karkee, M., & Zhang, Q. (2020). Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. *Computers and Electronics in Agriculture, 177*, 105687.
45. Lehnert, C., English, A., Mccool, C., Tow, A. W., & Perez, T. (2017). Autonomous sweet pepper harvesting for protected cropping systems. *IEEE Robotics and Automation Letters, 2*(2), 872–879.
46. Kwak, J., Lee, S., Baek, J., & Chu, B. (2022). Autonomous UAV target tracking and safe landing on a leveling mobile platform. *International Journal of Precision Engineering and Manufacturing, 23*(3), 305–317.
47. Park, J., An, B., Kwon, O., Yi, H., & Kim, C. (2022). User intention based intuitive mobile platform control: Application to a patient transfer robot. *International Journal of Precision Engineering and Manufacturing, 23*(6), 653–666.
48. Lee, D. G., Baek, D., Kim, H., Kim, J., & Kwon, D. (2023). Learning-based discrete hysteresis classifier using wire tension and compensator for flexible endoscopic surgery robots. *International Journal of Precision Engineering and Manufacturing, 24*(1), 83–94.
49. Yuan, L. H., Zhao, J. C., Li, W. H., & Hou, J. (2023). Improved informed-RRT* based path planning and trajectory optimization for mobile robots. *International Journal of Precision Engineering and Manufacturing, 24*(3), 435–446.
50. Jun, J., Kim, J., Seol, J., & Son, H. I. (2021). Towards an efficient tomato harvesting robot: 3D perception, manipulation, and end-effector. *IEEE access, 9*, 17631–17640.
51. Kang, H., Zhou, H., & Chen, C. (2020). Visual perception and modeling for autonomous apple harvesting. *IEEE Access, 8*, 62151–62163.
52. Lehnert, C., Mccool, C., Sa, I., & Perez, T. (2020). Performance improvements of a sweet pepper harvesting robot in protected cropping environments. *Journal of Field Robotics, 37*(7), 1197–1223.
53. Xiong, Y., Ge, Y., & From, P. J. (2020). An obstacle separation method for robotic picking of fruits in clusters. *Computers and Electronics in Agriculture, 175*, 105397.
54. Gao, J., Zhang, F., Zhang, J., Yuan, T., Yin, J., Guo, H., & Yang, C. (2022). Development and evaluation of a pneumatic finger-like end-effector for cherry tomato harvesting robot in greenhouse. *Computers and Electronics in Agriculture, 197*, 106879.
55. Lin, G., Tang, Y., Zou, X., & Liu, J. (2019). Guava detection and pose estimation using a low-cost RGB-D sensor in the field. *Sensors, 19*(2), 428.
56. Zhang, K., Lammers, K., Chu, P., Li, Z., & Lu, R. (2021). System design and control of an apple harvesting robot. *Mechatronics, 79*, 102644.
57. Liu, T. H., Ehsani, R., Toudeshki, A., Zou, X. J., & Wang, H. J. (2018). Detection of citrus fruit and tree trunks in natural environments using a multi-elliptical boundary model. *Computers in Industry, 99*, 9–16.
58. Liu, J., Yuan, Y., Zhou, Y., Zhu, X., & Syed, T. N. (2018). Experiments and analysis of close-shot identification of on-branch citrus fruit with realsense. *Sensors, 18*(5), 1510.
59. Qureshi, W. S., Payne, A., Walsh, K. B., Linker, R., Cohen, O., & Dailey, M. N. (2017). Machine vision for counting fruit on mango tree canopies. *Precision Agriculture, 18*, 224–244.
60. Faisal, M., Albogamy, F., Elgibreen, H., Algabri, M., & Alqershi, F. A. (2020). Deep learning and computer vision for estimating date fruits type, maturity level, and weight. *IEEE Access, 8*, 206770–206782.
61. Bresilla, K., Perulli, G. D., Boini, A., Morandi, B., Corelli Grappadelli, L., & Manfrini, L. (2019). Single-shot convolution neural networks for real-time fruit detection within the tree. *Frontiers in plant science, 10*, 611.
62. Pourdarbani, R., Sabzi, S., Hernández-Hernández, M., Hernández-Hernández, J. L., García-Mateos, G., Kalantari, D., & Molina-Martínez, J. M. (2019). Comparison of different classifiers and the majority voting rule for the detection of plum fruits in garden conditions. *Remote sensing, 11*(21), 2546.
63. Zahid, A., Mahmud, M. S., & He, L. (2021). Technological advancements towards developing a robotic pruner for apple trees: A review. *Computers and Electronics in Agriculture, 189*, 106383.
64. Son, J., Kang, H. Y. A., & Kang, S. H. (2023). A review on robust control of robot manipulators for future manufacturing. *International Journal of Precision Engineering and Manufacturing, 24*(6), 1083–1102.
65. Bae, J., Moon, Y., Park, E., Kim, J., Jin, S., & Seo, T. (2022). Cooperative underwater vehicle-manipulator operation using redundant resolution method. *International Journal of Precision Engineering and Manufacturing, 23*(9), 1003–1017.
66. Levin, M., & Degani, A. (2019). A conceptual framework and optimization for a task-based modular harvesting manipulator. *Computers and Electronics in Agriculture, 166*, 104987.
67. Navas, E., Fernández, R., Sepúlveda, D., & Armada, M. (2021). Soft grippers for automatic crop harvesting: A review. *Sensors, 21*(8), 2689.
68. Zhang, B., Xie, Y., Zhou, J., Wang, K., & Zhang, Z. (2020). State-of-the-art robotic grippers, grasping and control strategies, as well as their applications in agricultural robots: A review. *Computers and Electronics in Agriculture, 177*, 105694.
69. Rachmawati, E., Supriana, I., Khodra, M. L., & Firdaus, F. (2022). Integrating semantic features in fruit recognition based on perceptual color and semantic template. *Information Processing in Agriculture, 9*(2), 316–334.
70. Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., & Zou, X. (2020). Recognition and localization methods for vision-based fruit picking robots: A review. *Frontiers in Plant Science, 11*, 510.
71. Xiong, Y., Ge, Y., Grimstad, L., & From, P. J. (2020). An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. *Journal of Field Robotics, 37*(2), 202–224.
72. Edan, Y., Rogozin, D., Flash, T., & Miles, G. E. (2000). Robotic melon harvesting. *IEEE Transactions on Robotics and Automation, 16*(6), 831–835.
73. Ji, W., Zhao, D., Cheng, F., Xu, B., Zhang, Y., & Wang, J. (2012). Automatic recognition vision system guided for apple harvesting robot. *Computers & Electrical Engineering, 38*(5), 1186–1195.
74. Wang, C., Tang, Y., Zou, X., Luo, L., & Chen, X. (2017). Recognition and Matching of Clustered Mature Litchi Fruits Using Binocular Charge-Coupled Device (CCD) Color Cameras. *Sensors, 17*(11), 2564.
75. Arad, B., Kurtser, P., Barnea, E., Harel, B., Edan, Y., & Ben-Shahar, O. (2019). Controlled lighting and illumination-independent

target detection for real-time cost-efficient applicationsl. The case study of sweet pepper robotic harvesting. *Sensors, 19*(6), 1390.

76. Okamoto, H., & Lee, W. S. (2009). Green citrus detection using hyperspectral imaging. *Computers and electronics in agriculture, 66*(2), 201–208.
77. Wendel, A., Underwood, J., & Walsh, K. (2018). Maturity estimation of mangoes using hyperspectral imaging from a ground based mobile platform. *Computers and Electronics in Agriculture, 155*, 298–313.
78. Fatchurrahman, D., Amodio, M. L., & Chiara, M. (2020). Early discrimination of mature-and immature-green tomatoes (*Solanum lycopersicum* L.) using fluorescence imaging method. *Postharvest Biology and Technology, 169*, 111287.
79. Feng, J., Zeng, L., & He, L. (2019). Apple fruit recognition algorithm based on multi-spectral dynamic image analysis. *Sensors, 19*(4), 949.
80. Li, J., Zhang, R., Li, J., Wang, Z., Zhang, H., Zhan, B., & Jiang, Y. (2019). Detection of early decayed oranges based on multispectral principal component image combining both bi-dimensional empirical mode decomposition and watershed segmentation method. *Postharvest Biology and Technology, 158*, 110986.
81. Gan, H., Lee, W. S., Alchanatis, V., Ehsani, R., & Schueller, J. K. (2018). Immature green citrus fruit detection using color and thermal images. *Computers and Electronics in Agriculture, 152*, 117–125.
82. Osroosh, Y., & Peters, R. T. (2019). Detecting fruit surface wetness using a custom-built low-resolution thermal-RGB imager. *Computers and Electronics in Agriculture, 157*, 509–517.
83. Gan, H., Lee, W. S., Alchanatis, V., & Abd-Elrahman, A. (2020). Active thermal imaging for immature citrus fruit detection. *Biosystems Engineering, 198*, 291–303.
84. Iqbal, Z., Khan, M. A., Sharif, M., & Shah, J. H. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Computers and electronics in agriculture, 153*, 12–32.
85. Hameed, K., Chai, D., & Rassau, A. (2018). A comprehensive review of fruit and vegetable classification techniques. *Image and Vision Computing, 80*, 24–44.
86. Zhang, L., Yang, Q., Xun, Y., Chen, X., Ren, Y., Yuan, T., Tan, Y., & Li, W. (2007). Recognition of greenhouse cucumber fruit using computer vision. *New Zealand Journal of Agricultural Research, 50*(5), 1293–1298.
87. Fan, P., Lang, G., Yan, B., Lei, X., Guo, P., Liu, Z., & Yang, F. (2021). A method of segmenting apples based on gray-centered RGB color space. *Remote Sensing, 13*(6), 1211.
88. Jidong, L., De-An, Z., Wei, J., & Shihong, D. (2016). Recognition of apple fruit in natural environment. *Optik, 127*(3), 1354–1362.
89. Zhao, C., Lee, W. S., & He, D. (2016). Immature green citrus detection based on colour feature and sum of absolute transformed difference (SATD) using colour images in the citrus grove. *Computers and Electronics in Agriculture, 124*, 243–253.
90. Tan, K., Lee, W. S., Gan, H., & Wang, S. (2018). Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes. *Biosystems engineering, 176*, 59–72.
91. Yang, X., Zhang, R., Zhai, Z., Pang, Y., & Jin, Z. (2019). Machine learning for cultivar classification of apricots (*Prunus armeniaca* L.) based on shape features. *Scientia Horticulturae, 256*, 108524.
92. Lin, G., Tang, Y., Zou, X., Xiong, J., et al. (2020). Fruit detection in natural environment using partial shape matching and probabilistic Hough transform. *Precision Agriculture, 21*(1), 160–177.
93. Yamamoto, K., Guo, W., & Yoshioka, Y. (2014). On plant detection of intact tomato fruits using image analysis and machine learning methods. *Sensors, 14*(7), 12191–12206.
94. Li, H., Lee, W. S., & Wang, K. (2016). Immature green citrus fruit detection and counting based on fast normalized cross correlation (FNCC) using natural outdoor colour images. *Precision Agriculture, 17*(6), 678–697.
95. Zhang, C., Zou, K., & Pan, Y. (2020). A method of apple image segmentation based on color-texture fusion feature and machine learning. *Agronomy, 10*(7), 972.
96. Lin, G., & Zou, X. (2018). Citrus segmentation for automatic harvester combined with adaboost classifier and Leung-Malik filter bank. *IFAC-PapersOnLine, 51*(17), 379–383.
97. Wu, G., Zhu, Q., Huang, M., Guo, Y., & Qin, J. (2019). Automatic recognition of juicy peaches on trees based on 3D contour features and colour data. *Biosystems Engineering, 188*, 1–13.
98. Wu, G., Li, B., Zhu, Q., Huang, M., & Guo, Y. (2020). Using color and 3D geometry features to segment fruit point cloud and improve fruit recognition accuracy. *Computers and electronics in agriculture, 174*, 105475.
99. Ren, S., Zhang, Y., Sakao, T., Liu, Y., & Cai, R. (2022). An advanced operation mode with product-service system using lifecycle big data and deep learning. *International Journal of Precision Engineering and Manufacturing-Green Technology, 9*(1), 287–303.
100. Zheng, C., Li, W., Li, W., Xu, K., Peng, L., & Cha, S. W. (2022). A deep reinforcement learning-based energy management strategy for fuel cell hybrid buses. *International Journal of Precision Engineering and Manufacturing-Green Technology, 9*(3), 885–897.
101. He, K., Gkioxari, G., Dollár, P. (2017) Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*.
102. Huang, W. W., Gao, X. D., Huang, Y. H., & Zuang, Y. (2023). Improved convolutional neural network for laser welding defect prediction. *International Journal of Precision Engineering and Manufacturing, 24*(1), 33–41.
103. Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., & Zheng, Y. (2020). Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Computers and Electronics in Agriculture, 172*, 105380.
104. Parvathi, S., & Selvi, S. T. (2021). Detection of maturity stages of coconuts in complex background using Faster R-CNN model. *Biosystems engineering, 202*, 119–132.
105. Huang, G., Liu, Z., Van Der Maaten, L. (2017) Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
106. He, K., Zhang, X., Ren, S., Sun, J. (2016) Identity mappings in deep residual networks. In *Proceedings of the European conference on computer vision*. Springer.
107. Liu, X., Zhao, D., Jia, W., Ji, W., Ruan, C., & Sun, Y. (2019). Cucumber fruits detection in greenhouses based on instance segmentation. *IEEE Access, 7*, 139635–139642.
108. Shen, L., Su, J., Huang, R., Quan, W., Song, Y., Fang, Y., & Su, B. (2022). Fusing attention mechanism with Mask R-CNN for instance segmentation of grape cluster in the field. *Frontiers in plant science, 13*, 934450.
109. Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R., & Cui, Y. (2019). Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. *IEEE Access, 8*, 2327–2336.
110. Simonyan, K., Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:14091556.
111. Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture, 163*, 104846.

112. Hu, C., Liu, X., Pan, Z., et al. (2019). Automatic detection of single ripe tomato on plant combining faster R-CNN and intuitionistic fuzzy set. *IEEE Access, 7*, 154683–154696.
113. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
114. Redmon, J., Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition*.
115. Redmon, J., Farhadi, A. (2018). *Yolov3: An incremental improvement*. arXiv preprint arXiv:180402767
116. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M. (2020) *Yolov4: Optimal speed and accuracy of object detection*. arXiv preprint arXiv:200410934
117. Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture, 157*, 417–426.
118. Yan, B., Fan, P., Lei, X., Liu, Z., & Yang, F. (2021). A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sensing, 13*(9), 1619.
119. Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
120. Lv, J., Xu, H., Han, Y., Lu, W., Xu, L., Rong, H., Yang, B., Zou, L., & Ma, Z. (2022). A visual identification method for the apple growth forms in the orchard. *Computers and Electronics in Agriculture, 197*, 106954.
121. Tan, M., Pang, R., Le ,Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.
122. Chen, W., Lu, S., Liu, B., Chen, M., Li, G., & Qian, T. (2022). CitrusYOLO: A algorithm for citrus detection under orchard environment based on YOLOv4. *Multimedia Tools and Applications, 81*(22), 31363–31389.
123. Yang, L., Zhang, R. Y., Li, L., & Xie, X. (2021). Simam: A simple, parameter-free attention module for convolutional neural networks. In *Proceedings of the International conference on machine learning, PMLR.*
124. Chen, D., Tang, J., Xi, H., & Zhao, X. (2021). Image recognition of modern agricultural fruit maturity based on internet of things. *Traitement du Signal, 38*(4), 1237.
125. Chen, J., Wang, Z., & Wu, J. (2021). An improved Yolov3 based on dual path network for cherry tomatoes detection. *Journal of Food Process Engineering, 44*(10), e13803.
126. Li, H., Li, C., Li, G., & Chen, L. (2021). A real-time table grape detection method based on improved YOLOv4-tiny network in complex background. *Biosystems Engineering, 212*, 347–359.
127. Bodla N., Singh B., Chellappa R., & Davis, L. S. (2017). Soft-NMS--improving object detection with one line of code. In *Proceedings of the IEEE international conference on computer vision*.
128. Wang, F., Sun, Z., Chen, Y., et al. (2022). Xiaomila green pepper target detection method under complex environment based on improved YOLOv5s. *Agronomy, 12*(6), 1477.
129. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.
130. Zhou, Z., Song, Z., Fu, L., Gao, F., Li, R., & Cui, Y. (2020). Real-time kiwifruit detection in orchard using deep learning on Android™ smartphones for yield estimation. *Computers and Electronics in Agriculture, 179*, 105856.
131. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
132. Li, D., Sun, X., Elkhouchlaa, H., Jia, Y., Yao, Z., Lin, P., Li, J., & Lu, H. (2021). Fast detection and location of longan fruits using UAV images. *Computers and Electronics in Agriculture, 190*, 106465.
133. Wang, Y., Chen, Y., & Wang, D. (2022). Recognition of multimodal fusion images with irregular interference. *PeerJ Computer Science, 8*, e1018.
134. Mehta, S., & Burks, T. (2014). Vision-based control of robotic manipulator for citrus harvesting. *Computers and Electronics in Agriculture, 102*, 146–158.
135. Xiong, J., Liu, Z., Lin, R., Bu, R., He, Z., Yang, Z., & Liang, C. (2018). Green grape detection and picking-point calculation in a night-time natural environment using a charge-coupled device (CCD) vision sensor with artificial illumination. *Sensors, 18*(4), 969.
136. Mehta, S. S., Ton, C., Asundi, S., & Burks, T. F. (2017). Multiple camera fruit localization using a particle filter. *Computers and Electronics in Agriculture, 142*, 139–154.
137. Díaz, C. A., Pérez, D. S., Miatello, H., & Bromberg, F. (2018). Grapevine buds detection and localization in 3D space based on structure from Motion and 2D image classification. *Computers in Industry, 99*, 303–312.
138. Nguyen, T. T., Vandevoorde, K., Wouters, N., Kayacan, E., De Baerdemaeker, J. G., & Saeys, W. (2016). Detection of red and bicoloured apples on tree with an RGB-D camera. *Biosystems Engineering, 146*, 33–44.
139. Hou, C., Zhang, X., Tang, Y., Zhuang, J., Tan, Z., Huang, H., & Luo, S. (2022). Detection and localization of citrus fruit based on improved You Only Look Once v5s and binocular vision in the orchard. *Frontiers in Plant Science, 13*, 972445.
140. Li, T., Feng, Q., Qiu, Q., Xie, F., & Zhao, C. (2022). Occluded apple fruit detection and localization with a frustum-based point-cloud-processing approach for robotic harvesting. *Remote Sensing, 14*(3), 482.
141. Wu, F., Duan, J., Ai, P., Chen, Z., Yang, Z., & Zou, X. (2022). Rachis detection and three-dimensional localization of cut off point for vision-based banana robot. *Computers and Electronics in Agriculture, 198*, 107079.
142. Tian, Y., Duan, H., Luo, R., Zhang, Y., Jia, W., Lian, J., & Li, C. (2019). Fast recognition and location of target fruit based on depth information. *IEEE Access, 7*, 170553–170563.
143. Li, J., Tang, Y., Zou, X., Lin, G., & Wang, H. (2020). Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots. *IEEE Access, 8*, 117746–117758.
144. SepúLveda, D., Fernández, R., Navas, E., Armada, M., & Gonzalez-De-Santos, P. (2020). Robotic aubergine harvesting using dual-arm manipulation. *IEEE Access, 8*, 121889–121904.
145. Costa, J. M., Vaz, M., Escalona, J., Egipto, R., Lopes, C., Medrano, H., & Chaves, M. M. (2016). Modern viticulture in southern Europe: Vulnerabilities and strategies for adaptation to water scarcity. *Agricultural Water Management, 164*, 5–18.
146. Gongal, A., Amatya, S., Karkee, M., & Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. *Computers and Electronics in Agriculture, 116*, 8–19.
147. Giancola, S., Valenti, M., & Sala, R. (2018). *A survey on 3D cameras: Metrological comparison of time-of-flight, structured-light and active stereoscopy technologies*. Springer.
148. Wang, C., Luo, T., Zhao, L., Tang, Y., & Zou, X. (2019). Window zooming–based localization algorithm of fruit and vegetable for harvesting robot. *IEEE Access, 7*, 103639–103649.
149. Liu, T. H., Nie, X. N., Wu, J. M., Zhang, D., Liu, W., Cheng, Y. F., Qiu, J., & Qi, L. (2023). Pineapple (*Ananas comosus*) fruit detection and localization in natural environment based on

binocular stereo vision and improved YOLOv3 model. *Precision Agriculture, 24*(1), 139–160.

150. Xiong, J., He, Z., Lin, R., Liu, Z., Bu, R., Yang, Z., Peng, H., & Zou, X. (2018). Visual positioning technology of picking robots for dynamic litchi clusters with disturbance. *Computers and Electronics in Agriculture, 151*, 226–237.
151. Wang, M.-S. (2018). Eye to hand calibration using ANFIS for stereo vision-based object manipulation system. *Microsystem Technologies, 24*, 305–317.
152. Wang, X., Kang, H., & Zhou, H. (2022). Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. *Computers and Electronics in Agriculture, 193*, 106716.
153. Hutchinson, S., Hager, G. D., & Corke, P. I. (1996). A tutorial on visual servo control. *IEEE transactions on robotics and automation, 12*(5), 651–670.
154. Chaumette, F., & Hutchinson, S. (2006). Visual servo control. I. Basic approaches. *IEEE Robotics & Automation Magazine, 13*(4), 82–90.
155. Corke, P. I., Hager, G. D. (1998). Vision-based robot control. In *Control problems in robotics and automation*. (pp. 177–92). Springer.
156. Ling, X., Zhao, Y., Gong, L., Liu, C., & Wang, T. (2019). Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision. *Robotics and Autonomous Systems, 114*, 134–143.
157. Chen, W., Xu, T., Liu, J., Wang, M., & Zhao, D. (2019). Picking robot visual servo control based on modified fuzzy neural network sliding mode algorithms. *Electronics, 8*(6), 605.
158. Silwal, A., Davidson, J. R., Karkee, M., Mo, C., Zhang, Q., & Lewis, K. (2017). Design, integration, and field evaluation of a robotic apple harvester. *Journal of Field Robotics, 34*(6), 1140–1159.
159. Barth, R., Hemming, J., & Van Henten, E. J. (2016). Design of an eye-in-hand sensing and servo control framework for harvesting robotics in dense vegetation. *Biosystems Engineering, 146*, 71–84.
160. De-An, Z., Jidong, L., Wei, J., Ying, Z., & Yu, C. (2011). Design and control of an apple harvesting robot. *Biosystems engineering, 110*(2), 112–122.
161. Hussein, M. (2015). A review on vision-based control of flexible manipulators. *Advanced Robotics, 29*(24), 1575–1585.
162. Mehta, S., Mackunis, W., & Burks, T. (2014). Nonlinear robust visual servo control for robotic citrus harvesting. *IFAC Proceedings Volumes, 47*(3), 8110–8115.
163. Mehta, S. S., Mackunis, W., & Burks, T. F. (2016). Robust visual servo control in the presence of fruit motion for robotic citrus harvesting. *Computers and Electronics in Agriculture, 123*, 362–375.
164. Yu, X., Fan, Z., & Wang, X. (2021). A lab-customized autonomous humanoid apple harvesting robot. *Computers & Electrical Engineering, 96*, 107459.
165. Shirai, Y., & Inoue, H. (1973). Guiding a robot by visual feedback in assembling tasks. *Pattern recognition, 5*(2), 99–108.
166. Sun, X., Zhu, X., Wang, P. (2018). A review of robot control with visual servoing. In *Proceedings of the 2018 IEEE 8th annual international conference on CYBER Technology in automation, control, and intelligent systems (CYBER)*.
167. Feng, Q., Zou, W., Fan, P., Zhang, C., & Wang, X. (2018). Design and test of robotic harvesting system for cherry tomato. *International Journal of Agricultural and Biological Engineering, 11*(1), 96–100.
168. Nguyen, T. T., Kayacan, E., De Baedemaeker, J., & Saeys, W. (2013). Task and motion planning for apple harvesting robot. *IFAC Proceedings Volumes, 46*(18), 247–252.
169. Yeshmukhametov, A., Koganezawa, K., & Yamamoto, Y. (2022). Development of continuum robot arm and gripper for harvesting cherry tomatoes. *Applied Sciences, 12*(14), 6922.
170. Bac, C. W., Van Henten, E. J., Hemming, J., & Edan, Y. (2014). Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *Journal of Field Robotics, 31*(6), 888–911.
171. Rong, J., Wang, P., Yang, Q., & Huang, H. (2021). A field-tested harvesting robot for oyster mushroom in greenhouse. *Agronomy, 11*(6), 1210.
172. Wang, X., Kang, H., Zhou, H., Au, W., Wang, M. Y., & Chen, C. (2023). Development and evaluation of a robust soft robotic gripper for apple harvesting. *Computers and Electronics in Agriculture, 204*, 107552.
173. Li, S., Li, D., Zhang, C., & Xie, M. (2020). RGB-D Image Processing Algorithm for Target Recognition and Pose Estimation of Visual Servo System. *Sensors, 20*(2), 430.
174. Zubler, A. V., & Yoon, J.-Y. (2020). Proximal methods for plant stress detection using optical sensors and machine learning. *Biosensors, 10*(12), 193.
175. Lu, H., Li, Y., Uemura, T., Kim, H., & Serikawa, S. (2018). Low illumination underwater light field images reconstruction using deep convolutional neural networks. *Future Generation Computer Systems, 82*, 142–148.
176. Hua, X., Li, H., Zeng, J., Han, C., Chen, T., Tang, L., & Luo, Y. (2023). A review of target recognition technology for fruit picking robots: from digital image processing to deep learning. *Applied Sciences, 13*(7), 4160.
177. Barnett, J., Duke, M., Au, C. K., & Lim, S. H. (2020). Work distribution of multiple Cartesian robot arms for kiwifruit harvesting. *Computers and Electronics in Agriculture, 169*, 105202.
178. Chen, Z., Wu, R., Lin, Y., Li, C., Chen, S., Yuan, Z., & Zou, X. (2022). Plant disease recognition model based on improved YOLOv5. *Agronomy, 12*(2), 365.
179. Lu, Z., Zhao, M., Luo, J., Wang, G., & Wang, D. (2021). Design of a winter-jujube grading robot based on machine vision. *Computers and Electronics in Agriculture, 186*, 106170.
180. Apolo-Apolo, O. E., Martínez-Guanter, J., Egea, G., Raja, P., & Pérez-Ruiz, M. (2020). Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. *European Journal of Agronomy, 115*, 126030.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Jingfan Liu received the B.S. degree (2021) in Mechanical Engineering from Changsha University of Science and Technology, China. He is currently pursuing his M.S. degree in Mechanical Engineering at Wuhan University of Technology. His research interests include sensing and control of robotic systems.

Zhaobing Liu received his B.S. degree (2006) in Automation and the M.S. degree (2008) in Control Theory and Control Engineering from Northeastern University, China, and the Ph.D. degree (2014) in Mechanical Engineering from The University of Queensland, Australia. Now, he is an Associate Professor with Wuhan University of Technology. His current research interests include advanced manufacturing, robotics, system dynamics and control.