

Review

Research Progress and Development Trend of Visual Detection Methods for Selective Fruit Harvesting Robots

Wenbo Wang ^{1,*} , Chenshuo Li ¹, Yidan Xi ¹, Jinan Gu ¹ , Xinzhou Zhang ¹, Man Zhou ²  and Yuchun Peng ^{3,4,*}

¹ School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China

² School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

³ Department of Smart Agriculture and Engineering, Wenzhou Vocational College of Science and Technology, Wenzhou 325000, China

⁴ Wenzhou Key Laboratory of AI Agents for Agriculture, Wenzhou 325000, China

* Correspondence: wangwb@ujs.edu.cn (W.W.); pyc6559@hotmail.com (Y.P.)

Abstract

The rapid development of artificial intelligence technologies has promoted the emergence of Agriculture 4.0, where the machines participating in agricultural activities are made smart with the capacities of self-sensing, self-decision-making, and self-execution. As representative implementations of Agriculture 4.0, intelligent selective fruit harvesting robots demonstrate significant potential to alleviate labor-intensive demands in modern agriculture, where visual detection serves as the foundational component. However, the accurate detection of fruits remains a challenging issue due to the complex and unstructured nature of fruit orchards. This paper comprehensively reviews the recent progress in visual detection methods for selective fruit harvesting robots, covering cameras, traditional detection based on handcrafted feature methods, detection based on deep learning methods, and tree branch detection methods. Furthermore, the potential challenges and future trends of the visual detection system of selective fruit harvesting robots are critically discussed, facilitating a thorough comprehension of contemporary progress in this research area. The primary objective of this work is to highlight the pivotal role of visual perception in intelligent fruit harvesting robots.

Keywords: harvesting robots; selective fruit harvesting; visual detection; precision agriculture; Agriculture 4.0



Academic Editor: Dimitrios Kateris

Received: 10 July 2025

Revised: 6 August 2025

Accepted: 8 August 2025

Published: 10 August 2025

Citation: Wang, W.; Li, C.; Xi, Y.; Gu, J.; Zhang, X.; Zhou, M.; Peng, Y.

Research Progress and Development Trend of Visual Detection Methods for Selective Fruit Harvesting Robots.

Agronomy **2025**, *15*, 1926.

<https://doi.org/10.3390/agronomy15081926>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

([https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

[licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)).

1. Introduction

The evolution of technologies is constantly driving the upgrade of agricultural modes. The development roadmaps of the Agricultural Revolution and Industrial Revolution are presented in Figure 1 [1]. During the Agriculture 1.0 stage, the ancient and traditional farming practices heavily relied on manual work and animal power. The advent of Industry 1.0, characterized by mechanization, facilitated the application of agricultural machinery in sowing, irrigation, harvesting, and other processes, marking the emergence of Agriculture 2.0. Driven by the mass production modes brought by Industry 2.0 and digitalization by Industry 3.0, Agriculture 3.0 is enabled, which aims to introduce information and automation technologies for precision agriculture [2].

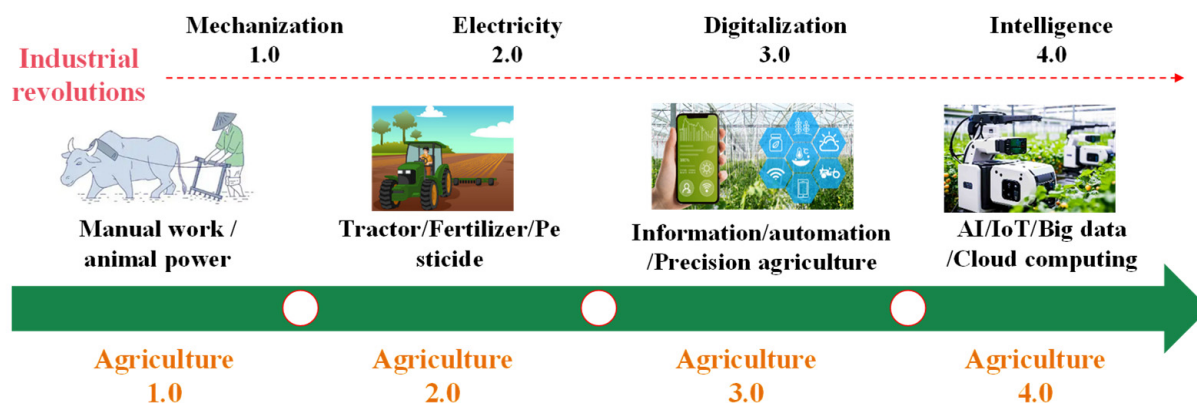


Figure 1. Development roadmap of Agricultural 4.0.

Recently, with the rapid development of Internet of Things [3], digital twin [4], Cyber Physical System [5], big data [6], and artificial intelligence [7,8], intelligent devices and smart manufacturing are enabled, and this industrial stage is called Industry 4.0. The integration of these advanced technologies and agriculture brings the chance to enter the era of Agriculture 4.0. During the Agriculture 4.0 environment, the real-time planting status can be monitored by sensors, and agricultural data can be analyzed on the virtual or cloud side automatically. Then, the production decisions can be made by themselves according to the sensed conditions, and their production parameters can be dynamically adjusted without consulting high-level managers.

During the Agriculture 4.0 era, the intelligent management of field crops is widely studied, such as irrigation [9,10], fertilization [11], disease classification [12], pesticide spraying [13], yield estimation [14], harvesting [15,16], and other aspects [17]. Intelligent robots are widely developed to counter labor shortages and provide an economically viable solution to rising labor costs. However, fruits are characterized by high vulnerability, disorderly growth, and asynchronous ripening, and the fruit growers need to finish a lot of picking work in a short time [18]. The development and application of intelligent fruit harvesting robots are essential for the fruit industry. Two principal harvesting approaches have been proposed, i.e., the mass harvesting method and selective harvesting method [19], as shown in Figure 2. Figure 2a shows a harvesting machine, which mainly applies vibration to the fruit tree to separate and collect the fruits. Nevertheless, it frequently causes damage to fruits and cannot work well for fresh fruit picking. In contrast, harvesting high-value fruits (e.g., apples, citrus, and tomatoes) still primarily relies on human labor, representing agriculture's labor-intensive and costly activity [20].



(a) Mass harvesting machine



(b) Selective harvesting robot

Figure 2. Fruit harvesting methods.

Figure 2b presents a selective harvesting robot, which picks each fruit at a time. The robotic system of a selective harvesting robot often consists of several sub-systems: the vision [21,22], end effector [23], and control [24]. The recognition of fruits is the premise for picking fruits by robots [25]. Advanced machine vision technologies are first applied to identify and localize the fruits [26,27]. After the fruit targets are detected, motion planning methods are used to guide the robotic arm to accurately reach the location of the fruits. Since each of these sub-systems is a complex research topic by itself, this paper mainly focuses on the visual detection methods for the selective fruit harvesting robots. Some review works have analyzed the robotic harvesting technologies in the agricultural area from discrete perspectives. Jia et al. [28] reviewed the apple harvesting robot in information technology. Khan et al. [29] reviewed the object detection methods in agricultural areas, covering crop monitoring, weed management, etc. Jin et al. [30] presented the development status and trend of agricultural robot technology. Wang et al. [31] discussed the main applications and progress of visual-based intelligent tea picking. Bai et al. [32] concentrated on the key navigation and guidance of intelligent robots in agriculture based on vision technologies. Xiao et al. [33] analyzed the main works in fruit detection methods based on traditional image processing and machine learning methods for smart harvesting robots. Hua et al. [34] discussed the key technologies in an apple harvesting robot for standardized orchards. Wu et al. [35] reviewed the key technologies for autonomous navigation in field agricultural machinery.

Nevertheless, existing visual detection methods of robotic fruit harvesting still face the challenges of reliable performance in real-life natural orchards. Few review papers have specifically focused on the visual detection methods for fruits. Considering these issues, it is necessary to answer the following questions: (1) What are the key components of fruit harvesting robots, and how does visual detection technology support the robots? (2) What are the current popular devices and algorithms in the field of vision-based recognition systems? (3) What are the main challenges and future trends of applying the cutting-edge vision-based fruit harvesting robots?

To answer these questions, this paper adopts a comprehensive methodological framework designed to analyze the advanced visual detection methods for selective fruit harvesting robots. The Web of Science (WoS) was selected as the sole bibliographic database for this study based on the following justifications: (1) its comprehensive interdisciplinary coverage and well-documented reputation for indexing high-quality, peer-reviewed scholarly publications [36]; (2) the use of a single authoritative database improves methodological reproducibility, reduces redundancy in retrieved records, and enhances the efficiency of the literature review process; and (3) the search was temporally restricted to publications between 2016 and 2025 to ensure the inclusion of contemporaneous and methodologically relevant studies. As shown in Figure 3, Section 2 introduces the background of fruit harvesting robots and visual detection methods based on relevant literature. Section 3 reviews the selected literature. Section 4 analyzes the main challenges and the future trends. Finally, Section 5 summarizes this study's principal conclusions.

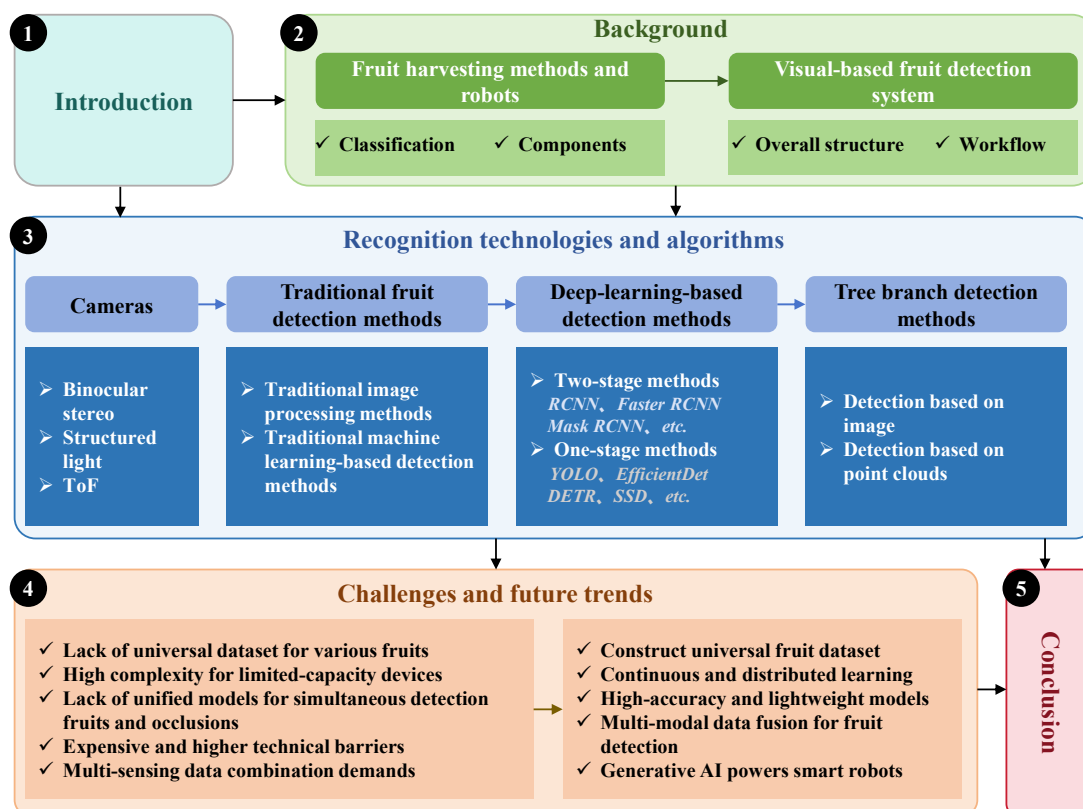


Figure 3. An overview of this review.

2. Background

2.1. Fruit Harvesting Methods and Robots

2.1.1. Fruit Harvesting Methods

To reduce the labor consumption of fruit harvesting, two main kinds of methods are often used, i.e., selective harvesting and mass harvesting [37].

Mass harvesting methods often use vibrational harvesting (i.e., catch and shake) and demonstrate significant potential for enhancing harvest efficiency and reducing production costs for tree fruit crops. Zhou et al. [38] analyzed the shaking-induced cherry fruit motion and damage; results showed that a shorter duration of high-level mechanical impacts correlates significantly with increased fruit damage. Sola-Guirado et al. [39] conducted a performance assessment of lateral canopy shakers incorporating a catch frame for orange harvesting in the juice production system. Results denoted that while operational, these systems require optimization to minimize both post-harvest ground fruit accumulation (5.9–10.4%) and mechanical damage to tree structures, particularly from trunk–catch frame interactions and branch–rod contacts. A high percentage of shaking-induced fruit and tree damage is the predominant limitation in the mechanical harvesting of fresh-market fruits [40]. In addition, the maturity of fruits is hard to maintain, so less mature fruits are inevitably harvested together with over-mature fruits. Therefore, this method is often used for harvesting fruits with relatively solid outer shells and is more suitable for the fruit industry, such as the juice industry.

Selective harvesting methods synergistically integrate the operational efficiency of mechanized systems with the selectivity of human labor. A robot is usually an integrated multi-disciplinary system that combines multiple functions [41], such as environment sensing, target detection and localization, motion planning and optimization, arm navigation, and robotic manipulation. As shown in Figure 4a, some robots are developed and applied for selective fruit harvesting. Levin et al. designed a peach harvesting robot with modular

reconfigurable joints [42]. Barnett et al. discussed a kiwifruit harvesting robot by employing multiple robot arms [43]. Zhang et al. proposed a robotic apple harvesting prototype combining synergistic mechatronic design and motion control [44]. Japer et al. designed an autonomous mobile robot collector for oil palm loose fruits [45]. Xiao et al. presented a citrus picking robot with a flexible hand claw and a target detection system [46].

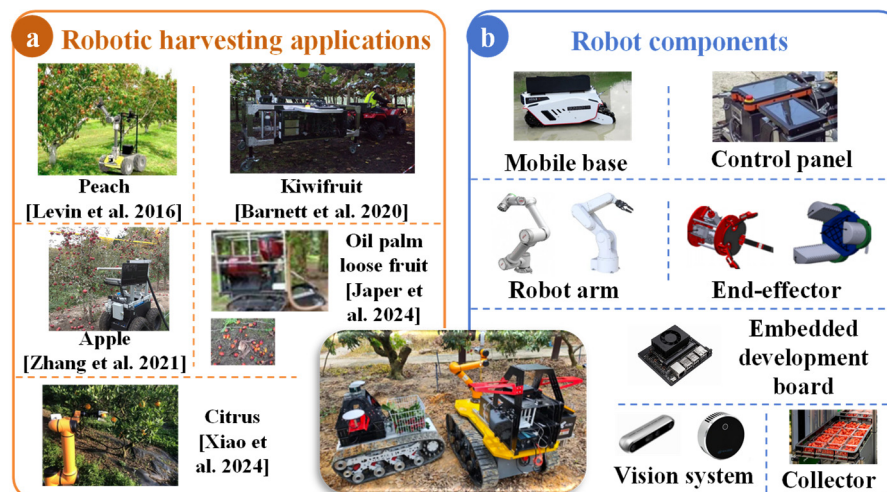


Figure 4. Applications and components for selective harvesting robots [42–46].

2.1.2. Fruit Harvesting Robot Components

The popular selective harvesting robots are composed of several main parts, i.e., mobile base, robot arm, control panel, end effector, embedded development board, machine vision system, and collector, as shown in Figure 4b.

- The mobile base bears the basic functions, such as positioning, navigation, and obstacle avoidance of the robot [47]. It is not only the integration point for various sensors, motors, and other equipment but also the key for the robot to achieve autonomous walking, obstacle avoidance, and path planning in natural orchards.
- The robot arm aims to approach the fruits while avoiding the obstacles (e.g., tree branches). The route optimization of the arm and collaboration between different arms are widely researched [48,49].
- The control panel is used to exchange information between the farmer or operator and the robot [50].
- The end effector or the gripper is set to detach and collect the targeted fruits [51], including a flexible three-fingered end effector for apple [52,53] and tomato [54] harvesting and a grip-and-cut end effector for grape clusters [55,56].
- The embedded development board is employed to run the analysis models (e.g., the visual detection model) on the edge side (the robot itself).
- The machine vision system is the eye for the robot, which can detect and localize fruits and obstacles [57,58]. It comprises hardware and software systems. The hardware system contains a camera, a server, information exchange devices, etc., and the software systems mainly consist of a visual-based fruit detection method.
- The collector is set to collect and temporarily store fruits.

2.2. Visual-Based Fruit Detection System

To achieve the autonomous operation of selective fruit harvesting robots, the visual detection of the fruits is the first basic step. It aims to construct a computational model that can autonomously fulfill visual detection tasks like human eyes [59,60]. Figure 5a presents the overall structure for a visual-based fruit detection system. As shown in Figure 5b, the

workflow of the method mainly consists of four procedures: (1) fruit image collection, (2) image data processing, (3) model learning, and (4) task execution.

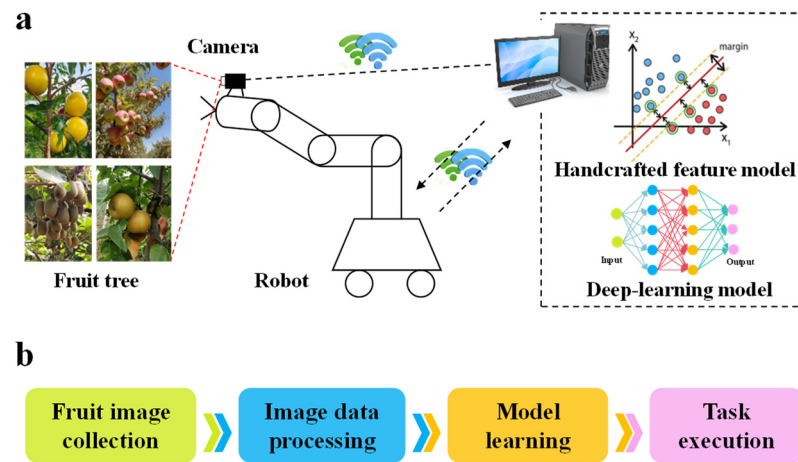


Figure 5. Overall structure for visual-based fruit detection system: (a) overall structure and (b) workflow.

(1) In the fruits image collection stage, the camera on the robot arm side collects the fruits image or videos. To ensure the diversity of samples, the shooting should be conducted under different lighting conditions (including front light, side light, and back light) and various angles so that the dataset can cover a variety of situations in natural scenes [61].

(2) The raw sensed fruit images are processed to construct datasets for model learning. The image annotation software (e.g., Labellmg) is used to mark the fruits in the picture. To solve the problem of small and single datasets, the data augmentation methods [62] are often implemented. These methods generate more diverse and richer training samples by performing various transformations and processing on the original images, including Brightness adjustment, panning, contrast transformation, Gaussian blur, and adding noise. In addition, several public datasets can be used, e.g., the date fruit dataset [63], citrus fruit dataset (CitDet) [64], etc.

(3) The model learning stage aims to train the fruit detection models on the server side. Two primary models are studied, i.e., hand-crafted feature model and deep learning model [65]. The former methods extract features (i.e., color, shape, and texture) from the target objects, followed by a machine learning-based classifier for fruit detection. The deep learning methods learn the final task objective directly from the original data, without relying on manually designed feature extractors, and they are now popular measures in the fruit detection [66]. The main works will be reviewed in Sections 3.2 and 3.3 in detail.

(4) The task execution stage uses the obtained visual detection model to fulfill the fruit detection targets and then transfers the detected fruit information to the control panel for further picking.

3. Recognition Technologies and Algorithms

3.1. Cameras

The camera is one of the most prevalent robotic sensors, enabling non-contact environmental perception through optical measurement. Based on the imaging principle, the camera used by fruit harvesting robots can be classified into three main kinds, i.e., binocular stereo camera, structured light cameras, and Time of Flight (ToF) cameras.

3.1.1. Binocular Stereo Cameras

The depth camera, based on binocular stereo vision, is similar to the human eye. Binocular stereo cameras, unlike depth cameras based on ToF or structured light principles, do not actively project light. It solely relies on capturing two images (either color RGB or grayscale) to compute depth and is therefore sometimes referred to as a passive binocular depth camera. Well-known products include the ZED 2K Stereo Camera by STEROLABS (San Francisco, CA, USA) and the BumbleBee by Point Grey (Vancouver, BC, Canada), as shown in Figure 6a.

As shown in Figure 6b, after binocular camera calibration and stereo rectification, stereo matching is performed to find the corresponding pixels P of the same spatial point in the left and right images [67]. Then, the spatial point P is projected into the image planes of the left and right cameras, forming two image points $p(u_L, v_L)$ and $p(u_R, v_R)$. By matching these two points, the horizontal position difference (disparity) is calculated. Using the disparity, camera focal length f , and baseline b , the 3D coordinates of the spatial point P can be calculated.

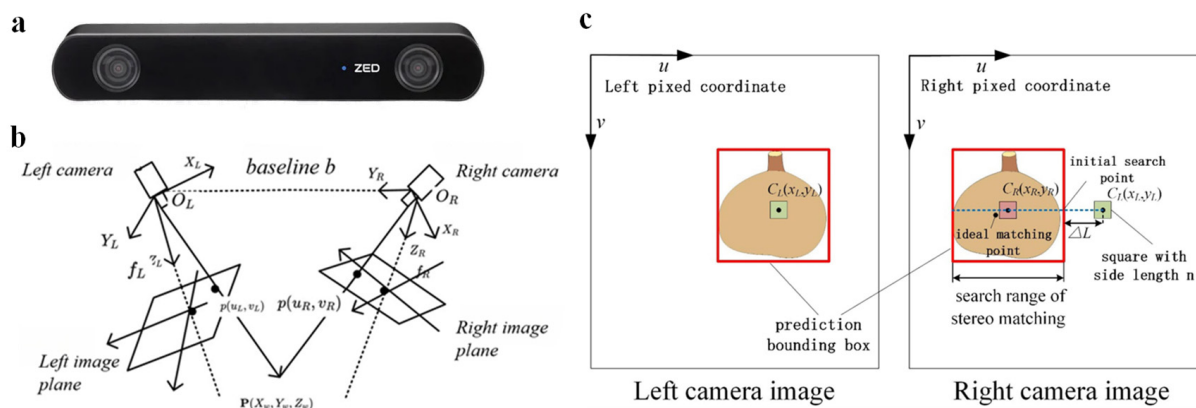


Figure 6. Binocular stereo cameras: (a) an example, (b) general principle [66], and (c) stereo matching [68].

The depth Z_W of the spatial point P can be calculated as follows:

$$Z_W = \frac{f \cdot b}{u_L - u_R} \quad (1)$$

In fruit harvesting robots, the procedures of binocular stereo vision based on deep learning typically involve the following steps [68].

(1) Binocular Calibration: The goal is to determine the geometric relationship and optical parameters between the two cameras, allowing the two cameras to precisely register the same scene.

(2) Stereo Rectification: The goal is to transform the images from the two cameras in such a way that the pixels in both images are aligned along the same horizontal line.

(3) Stereo Matching: The goal is to find the matching positions of the same object points in the left and right camera images and then compute the disparity, which is the horizontal offset between corresponding pixels in the two images [69], as shown in Figure 6c. First, the binocular cameras capture synchronized left and right images and send them to a target detection model (such as YOLO) to recognize and obtain the region of interest (ROI) of the target and center coordinates. Stereo matching is then performed within the ROI, using matching algorithms to compute the binocular disparity between corresponding points in the left and right images.

(4) Triangulation: Based on the stereo matching results (disparity), the object's position in 3D space is recovered. Using the intrinsic and extrinsic parameters of both cameras, along with the disparity information, the 3D coordinates of the object can be calculated.

(5) Get the Picking Point: The 3D coordinates obtained from the above steps are utilized for precise spatial localization of the target fruit. By utilizing these coordinates and the robot's kinematic model, the picking path can be planned.

Recent advances in the application of binocular stereo cameras for fruit detection and localization are as follows. The ZED2 binocular stereo camera platform was constructed to capture 30 pairs of binocular images of the *Camellia oleifera* fruits for obtaining their picking point [69]. Li et al. acquired the RGB (red, green, and blue) images with the binocular camera (BumbleBeeXB3) for apples detection [70], the geometric transformation and image enhancement was applied to annotate images, and stereo matching was performed via template matching between left and right images, with feature points extracted for disparity calculation and 3D coordinate reconstruction. Zhang et al. applied the zoom binocular camera to obtain images and then a novel depth estimation algorithm to perform stereo matching and triangulation [71].

The advantages of stereo or multi-view visual matching include the ability to acquire rich spatial depth information from multiple perspectives, improving the accuracy of 3D reconstruction, and object recognition. However, it involves high computational costs and complex algorithms. In addition, it is sensitive to environmental factors such as lighting changes, reflections, texture scarcity, and perspective differences, which can lead to matching errors or failures.

3.1.2. Structured Light Cameras

A structured light camera projects a specific light pattern onto the object's surface and decodes the reflected light to extract the object's 3D information. Figure 7 shows the examples and main principle of structured light cameras. Sinusoidal stripes are generated through computer programming and projected onto the object to be measured using a projection device. A Charge Coupled Device (CCD) camera captures the distortion of the stripes caused by the object. The deformed stripes are then demodulated to obtain the phase value of each pixel. The phase is then converted into the height of the entire field. Finally, by combining the camera's intrinsic parameters (focal length, principal point, and pixel size) and extrinsic parameters (the relative position between the camera and the projector), the height of each pixel is converted into three-dimensional coordinates.

Recent studies applied structured light cameras to obtain the raw images and depth information for fruit detection and localization. For example, Zhang et al. combined Structured Light Reflectance Imaging (SIRI) with the U-Net semantic segmentation model for early rot detection in oranges [72]. In parallel, Zhang et al. combined the industrial camera and Orbbec Astra S infrared structured light depth camera from Orbbec (Shenzhen, China) to obtain the RGB color image and RGB-D depth image, respectively, and then used the obtained images to detect and localize apples [73]. Li et al. applied the Intel RealSense D435i RGB-D camera from Intel Corporation (Santa Clara, CA, USA), which combined structured light-assisted stereo vision technology, used an infrared projector (IR Projector) and two infrared receivers (IR Receivers) to capture the depth and color images of the entire scene, and then a new deep learning instance segmentation method was developed 3D localization and grasp pose estimation of apples under occlusion conditions [74].

Structured light cameras can provide high-precision depth information rapidly, which is especially suitable for measuring objects at close range and static objects. However, they are susceptible to interference in strong light environments (such as outdoor sunlight), which can affect the quality of depth images. Additionally, they are only suit-

able for medium to short distances, and as the measurement distance increases, accuracy may decrease. They are also prone to motion artifacts in dynamic scenes or with fast-moving objects.

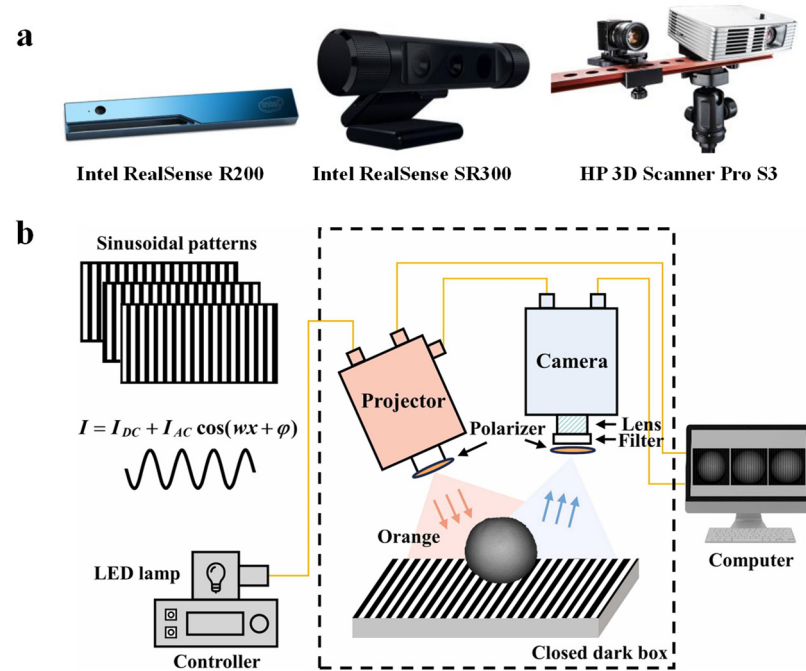


Figure 7. Structured light cameras: (a) examples and (b) general principle [72].

3.1.3. Time of Flight Cameras

ToF cameras calculate distance by measuring the travel time of light. The operational principle involves emission of modulated optical pulses toward target objects, with subsequent detection of the reflected waveforms by a photonic receiver. Depth information is derived through accurate measurement of the photon round-trip time interval between emission and detection events. Figure 8 shows some examples and principles of the ToF cameras.



Figure 8. Time of Flight cameras: (a) examples and (b) principle [75].

Recent advances in agricultural vision systems have demonstrated the effective application of ToF cameras in capturing depth images. Sun et al. applied the Microsoft Kinect V2 camera from Microsoft Corporation (Washington, DC, USA), which combines structured

light technology and ToF technology, to capture RGB images and depth images [76]. Legg et al. proposed a method for non-destructive maturity estimation of grapes using Time of Flight (Kinect Azure) and LiDAR (Intel L515) depth cameras [77]. Matthew Peebles et al. used a Microsoft Kinect v2 (ToF camera) to acquire depth and infrared images and a Basler ACE (RGB camera) from Basler AG (Arensburg, Germany) to capture color images [78].

A ToF camera does not require any external light source to scan the surrounding environment and can work normally even in conditions with little or no light. The measurement distance of ToF is relatively long, unaffected by surface gray scale and features, reaching up to a hundred meters. However, it is highly sensitive to environmental light intensity and interference and is basically unusable under strong outdoor light.

The wide application of depth cameras provides crucial three-dimensional spatial information (such as precise shape, size, and relative position) for fruit visual detection systems, significantly enhancing the foundation of perception capabilities. However, merely obtaining high-quality RGB-D data itself cannot directly complete the detection task. How to efficiently extract discriminative features from this heterogeneous (RGB + depth) or even high-dimensional (point cloud) data and thereby achieve accurate classification and localization is the core algorithmic challenge in fruit visual detection research. Therefore, the following two sections will systematically review and deeply analyze the key methods applied in fruit visual recognition. As shown in Figure 9, fruit detection methods can be classified into three main kinds, i.e., traditional image processing methods, traditional machine learning methods, and deep learning methods. The first two kinds of methods are mainly based on handcrafted features and will be discussed in detail in Section 3.2. Section 3.3 focuses on the deep learning techniques that have dominated this field in recent years.

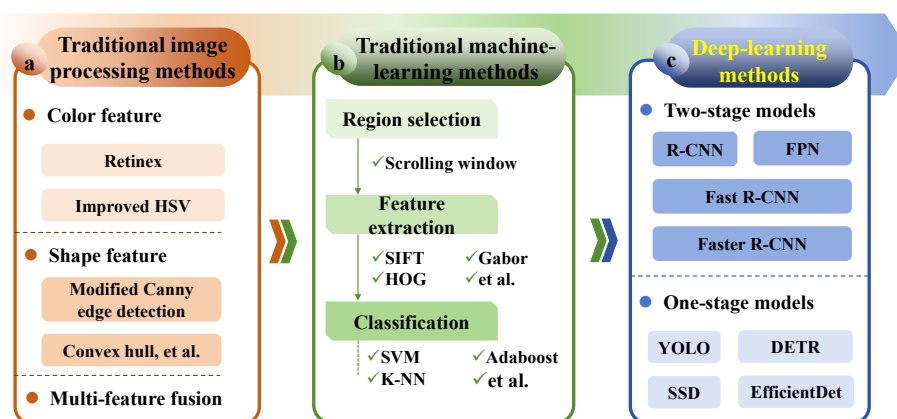


Figure 9. Development of fruit detection methods.

3.2. Traditional Fruit Detection Based on Handcrafted Features

Traditional fruit detection technologies mainly used basic image processing and machine learning methods, as shown in Figure 9a,b. In the early years, the fruits are detected based on their basic color and shape features since these features of fruits differ largely from those of leaves, branches, and the backgrounds [75]. Then, the machine learning methods are applied in the selective fruit detection area, which contains three main processes, i.e., region selection, feature extraction, and classification [79].

3.2.1. Traditional Image Processing Methods

Fruits typically exhibit stable and obvious features. The important features of the fruits include colors, shapes, textures, etc. These features can be utilized either individually or through multi-feature fusion to enable robust fruit identification in selective harvesting robotic systems. Substantial research efforts have been devoted to this field.

(a) Fruit detection based on color features.

Mature fruits normally have stable color characteristics, including color histogram, color set, and color coherence vector [79]. Color-based detection methods are primarily used when the target fruits (e.g., tomatoes, apples, and citrus) demonstrate significant chromatic contrast against their background environments. For example, to estimate the ripeness of tomatoes, a fuzzy rule-based classification approach was proposed based on the color, where two color depictions (red–green color difference and ratio) are derived from the extracted RGB color information [80]. A binocular vision system was developed for a lab-customized autonomous humanoid apple harvesting robot, incorporating color segmentation and morphological processing algorithms [81]. An automatic color determination procedure was presented for mango quality evaluation using a CCD camera and color standards [82]. The recognition of blueberry fruit and their maturity was fulfilled using Histogram-Oriented Gradients (HOGs) and color features in outdoor scenes [83]. An artificial olfactory system employing colorimetric sensor arrays was engineered for quantitative quality assessment of Chinese green tea [84]. To address the issue of difficulty in recognition due to the presence of many dark spaces, shadows, and low resolution in nighttime apple images, a Retinex algorithm based on guided filter according to the color feature of the image was presented to enhance nighttime images [85]. The proposed algorithm achieved superior image enhancement performance metrics compared to conventional histogram equalization methods and bilateral filtering-based Retinex algorithms.

In summary, the fruit detection methods based on color features are more applicable for objects whose color differs significantly from the background. However, they are susceptible to lighting variations and shadows, and most color-based techniques exhibit significant sensitivity to illumination variations and limited contrasts between fruits and leaves.

(b) Fruit detection methods based on shape features.

The geometric shape or edge feature is another type of key indicator for selective fruit detection [86], serving as more reliable discriminative features for fruit recognition under challenging lighting conditions. The distinct geometric disparity between the quasi-spherical morphology of fruits and the planar/linear structures of leaves and branches enables effective detection even when fruits lack distinctive chromatic separation from their background environment. Lin et al. introduced an innovative fruit detection methodology employing partial shape matching and probabilistic Hough transform [86]. Experiments validation across citrus, tomato, and pumpkin datasets confirmed the method's robust performance in detecting diverse fruit morphologies (including green/orange and circular/non-circular specimens) under natural conditions. Lv et al. proposed an apple image segmentation algorithm integrating a convex hull center priori with Markov adsorption chain, achieving exceptional performance metrics (IoU: 95.65%) with computational efficiency (average execution time: 1.02 s) [87]. For strawberry phenotyping, researchers implemented a computationally efficient algorithm leveraging geometric approximations using 'right kite' and 'simple kite' models [88]. The system attained classification accuracy of 94–97% with rapid processing speeds (<0.5 s per fruit). Lu et al. developed an immature citrus fruit detection method combining local binary pattern feature and hierarchical contour analysis [89], while another study addressed the challenge of bagged apple recognition under specular reflections through a block classification approach incorporating watershed segmentation of R-G edge-enhanced grayscale images [90].

These methods, based on shape features, are less affected by changes in light intensity. However, the randomness of fruit planting can affect their accuracy and speed. These methods are more suitable for orchards with better agricultural operations.

(c) Fruit detection based on multi-feature fusion.

The detection method based on a single kind of feature usually has limitations; thus, the detection methods by fusing two or more features are also studied. A novel automated tomato ripeness recognition algorithm was developed for robotic harvesting, integrating multiple features, feature analysis and selection, a weighted relevance vector machine classifier, and a bi-layer classification strategy [91]. For apple detection, a multimode detection approach based on color and shape features was proposed [92]. The method employed simple linear iterative clustering to segment orchard images into super-pixel blocks. The color feature extracted from blocks was used to determine candidate regions, and the HOG was adopted to describe the shape of fruits. In durian recognition systems, researchers implemented a linear discriminant analysis framework using both geometric features (area, perimeter, and circularity) for shape characterization and local binary patterns for texture analysis [93]. A machine vision algorithm employing an elliptical boundary model was developed for automated pomelo fruit detection [94]. The implemented computational framework comprised: (1) chromatic space conversion from RGB to YCbCr color space, followed by (2) implicit second-order polynomial fitting via ordinary least squares (OLS) regression to establish elliptical boundary models within the Cr-Cb chromatic subspace. This methodology facilitated precise segmentation of pomelo fruits according to five phenotypically distinct maturity stages.

Traditional image processing methods rely on manually designed feature extraction rules, such as color, shape, texture, etc. These methods perform well for specific environments and types of fruit, especially when conditions are relatively simple and predictable. Additionally, they do not require large amounts of training data and can process images directly based on predefined rules. However, they often exhibit poor robustness and a decline in accuracy when dealing with complex environments (such as lighting changes, partial occlusion of the target fruit, background interference, and image noise) and tasks that require recognition of fruits with different shapes. Especially for fruit-picking functions that demand high precision and robustness, traditional methods may fail to meet the practical requirements.

3.2.2. Traditional Machine Learning-Based Detection Methods

To improve the detection precision and robustness, machine learning technologies have been systematically integrated with traditional image processing procedures. As shown in Figure 9b, machine learning-based methods typically consist of three main parts: region selection, feature extraction, and classification [95]. The region selection module uses a fixed-sized sliding window to generate candidate target areas. The feature extraction module involves image processing techniques such as Gabor transform and Histogram of Oriented Gradients (HOGs). The classification module employs machine learning algorithms like Support Vector Machines (SVMs), Random Forest, K-Nearest Neighbor (K-NN) [96], and Decision Tree.

The K-means clustering is a popular machine learning algorithm, which is an unsupervised learning algorithm for partitioning data into K distinct clusters based on similarity. Fan et al. proposed a multi-feature patch-based apple image segmentation technique using the gray-centered RGB color space transformation [97]. This generalized K-means clustering approach comprehensively processed all image features (including specular highlights and shadow artifacts) to achieve optimized target extraction efficiency. For circular fruit localization, an overlapping detection methodology employing local maxima analysis was established [98]. This technique jointly uses lab color space conversion with K-means clustering, morphological contour refinement, and minimum edge-distance computation for centroid determination through an accelerated geometric algorithm. Additionally, a K-

means clustering-based recognition system was implemented for litchi detection, effectively separating target fruits from complex backgrounds (foliage and branches) [67].

SVM is a supervised learning algorithm for classification, whose core aim is to identify an optimal separating hyperplane that maximizes the margin between distinct classes within the feature space. Xu et al. presented a two-stage strawberry detection method for application in a strawberry harvesting robot, utilizing an HOG descriptor coupled with an SVM classifier [99]. Initially, potential strawberry regions were identified through HSV (hue, saturation, value) color segmentation, followed by the extraction of the HOG descriptor calculated using five ROI, which were subsequently fed into an HOG/SVM classifier for detecting the strawberries. Feng et al. proposed an innovative apple recognition algorithm combining multispectral dynamic imaging (MSX) technology, which was based on the pseudo-color and texture information of MSX images, and object classification using an SVM based on the extracted texture features [100]. Kumar et al. presented a microcontroller-based machine vision system employing an SVM classifier for tomato grading and sorting [101]. Peng et al. presented an identification method based on deep feature extraction and SVM, which can identify grape varieties rapidly and accurately [102].

The AdaBoost algorithm is an ensemble learning algorithm that combines multiple weak learners (slightly better than random guessing) to form a strong learner. It adaptively adjusts sample weights to focus on misclassified instances in each iteration. A tomato detection algorithm combining an AdaBoost classifier and color analysis was proposed and employed by the harvesting robot [103]. The results showed that over 96% of target tomatoes were correctly detected at a speed of about 10 fps. The positioning error of the robot end-point of less than 10 mm was achieved for large-scale direct positioning of the harvesting robot.

Bayesian algorithms are probabilistic methods based on Bayes' theorem; it has the ability to handle multiple classification tasks and provide better performance for fruit detection. For example, a real-time image analysis algorithm was developed for cherry classification by processing their digitized color images [104]. This method involved histogram analysis in both RGB and HSV color spaces, followed by the development of a hybrid Bayesian classification algorithm utilizing the components R and H.

In general, traditional machine learning-based fruit detection methods often perform better than traditional image processing methods. However, they still exhibit poor robustness, and performance tends to degrade when the size of the target varies greatly, or the shape of the target is complex and diverse. They also usually rely on manually designed feature extraction algorithms and are sensitive to abnormal data.

3.3. Fruit Detection Based on Deep Learning Methods

To overcome the limitations of traditional object detection methods, the development of deep learning technology in recent years has brought new solutions to fruit object detection. Deep learning algorithms, with their advantages of end-to-end recognition and automatic feature extraction, have significantly improved the detection accuracy and speed [105,106]. Deep learning models no longer rely on manually designed feature extractors or manual intervention in intermediate steps [107]. They can directly learn the final task objective from the raw data, greatly simplifying the entire object detection process and improving the efficiency of the model [108]. As shown in Figure 9c, the popular deep learning-based fruits detection methods can be classified into two main directions, i.e., two-stage models and one-stage models [109]. Two-stage fruit detection methods are usually divided into two stages: (1) set a Region Proposal Network (RPN) to generate candidate regions, followed by (2) classifying the candidate regions. One-stage fruit detection methods use a single neural network model to fulfill the object detection tasks.

3.3.1. Two-Stage Methods

The two-stage algorithm typically includes two categories: object detection and instance segmentation. The object detection algorithm first generates candidate object regions and performs object classification and localization, while the instance segmentation algorithm, based on object detection, further performs pixel-level segmentation, which not only distinguishes between various objects but also between instances of the same category. The common two-stage network frameworks include Region-based CNN (RCNN), Faster RCNN, Mask RCNN, etc.

RCNN is the earliest two-stage object detection algorithm, which used high-capacity CNNs for the first time to locate and identify objects using bottom-up region proposals [110]. RPN, a fully convolutional network that concurrently predicts object limits and objective scores at every position, is introduced by Faster RCNN. It shares full-image convolutional features with the detection network. End-to-end training of RPNs produces high-quality region proposals, which Fast RCNN uses for detection [111]. The structure of the Faster RCNN is shown in Figure 10.

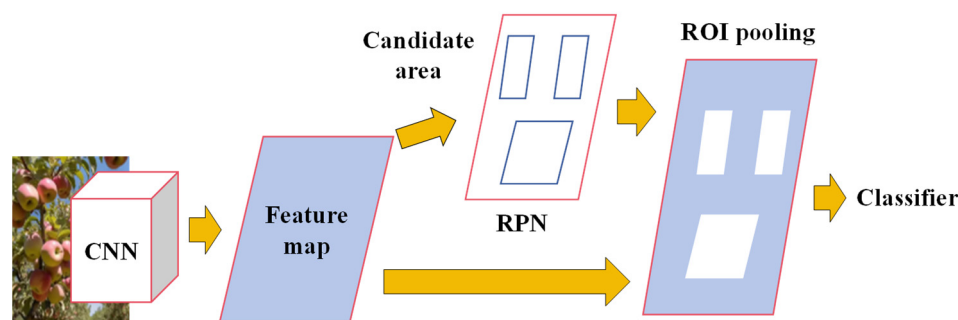


Figure 10. The structure of Faster RCNN.

Sa et al. adapted the Faster RCNN model, through transfer learning, for the fruit detection task employing images from two modalities, i.e., color (RGB) and Near-Infrared (NIR) [112]. In order to address the model's insensitivity to small tomato fruits, Wang et al. proposed a method for detecting young tomato fruits with a near-color background using an improved Faster RCNN with CBAM attention mechanism [113]. The method used ResNet50 as the backbone, and Feature Pyramid Network (FPN) to fuse high-level semantic features with low-level detailed features, achieving better performance than the original Faster RCNN. Based on improved Faster RCNN, Wan et al. proposed a deep learning framework for multi-class fruits detection [114]. The framework included image library creation, data augmentation, improved Faster RCNN model generation, and performance evaluation. In order to automatically recognize little fruits, Mai et al. proposed a new version of Faster RCNN with classifier fusion [115]. In the stage of proposal localization, three classifiers for object classification were learned using characteristics from three different levels. To estimate the citrus yields, a Faster RCNN model was constructed to detect the oranges from the images sensed by an unmanned aerial vehicle [116].

By adding a branch for object mask prediction in tandem with the current branch for bounding box identification, the Mask RCNN expands on Faster RCNN, enabling it to detect objects in an image efficiently and provide a high-quality segmentation mask for each instance [117]. Yu et al. introduced Mask-RCNN to improve the performance of machine vision for strawberry detection in harvesting robot systems [118]. Their framework employed ResNet-50 as the backbone network integrated with an FPN for multi-scale feature extraction, while utilizing an end-to-end trained RPN to generate region proposals across feature maps. Wang et al. advanced the field by introducing a refined Mask RCNN variant for precise apple instance segmentation [119]. Their architectural improvements

incorporated an attention mechanism within the backbone network to augment feature extraction capabilities, specifically implementing a hybrid module combining deformable convolution with transformer attention (key content only term). Lv et al. proposed a method that implemented a Mask RCNN augmented with PointRend for robust apple identification across varied growth morphologies in the orchard environments [120]. Wang et al. applied the dilated convolution to the res4b module of the Mask RCNN backbone network, ResNet, to realize the accurate identification and segmentation of waxberry [121].

In general, these two-stage methods can achieve high accuracy but require a rather long computation time. As a result, two-stage methods are unsuitable for real-time applications.

3.3.2. One-Stage Methods

One-stage object detection has relatively lower accuracy, especially in complex scenarios, but it generally has lower computational complexity, faster speed, and higher efficiency compared to two-stage methods. Therefore, in the development of practical fruit-picking robots, one-stage detection is more suitable for real-time applications due to its faster detection speed. Typical one-stage object detection algorithms include You Only Look Once (YOLO) series [122,123], Single Shot multiBox Detector (SSD) [124], Detection Transformers (DETRs), EfficientDet [125], etc.

(a) Fruit detection methods based on the YOLO series algorithm.

In fruit picking, the most commonly used one-stage object detection method is undoubtedly the YOLO series [126]. Figure 11 shows the framework of the YOLO series algorithms for object detection, which includes two main parts, i.e., the detection process and the training process [127]. During the detection process, after the origin image is input, the simple preprocessing methods are called to resize and divide the images. The convolutional neural network, which generally contains a convolutional layer, a pooling layer, and a fully connected layer, is set to obtain the prediction information. Then, the postprocessing process is applied to filter redundant information and, finally, to integrate the image and output. The modules of the training process share similar steps with the detection process. One different module is the network parameter adjustment, which optimizes the parameters of the network to keep the consistency between the network output and the real value.

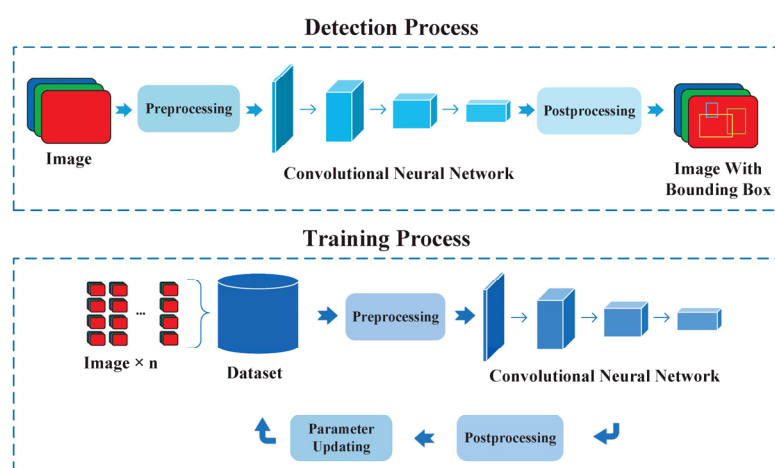


Figure 11. The framework of the YOLO series algorithms for object detection [127].

The inaugural version of YOLO was formally introduced and published by Joseph Redmon et al., seeking to establish a revolutionary single-stage model predicated on end-to-end training, real-time inference, and customizable optimization [128]. The YOLO treated each region as a potential candidate object for detection and directly predicted bounding

boxes and class probabilities from the entire image. This approach simplified the detection process, allowing for end-to-end optimization and achieving real-time performance. YOLO, as a unified object detection model, has become a pioneering work in the field of real-time object detection due to its simplicity, speed, and accuracy. To recognize oil palm loose fruits, an improved YOLO model was developed [129], which combined a densely connected neural network, swish activation function, multi-layer detection, and prior box optimization. Subsequent versions have continuously improved and optimized, further enhancing detection performance and generalization ability [130,131]. Table 1 shows a comparison between major fruit detection approaches based on YOLO series algorithms. The dataset, feature, their accuracy (mAP), and inference speed are presented for each model.

YOLOv3 uses Darknet-53 as the backbone network and extensively employs residual skip connections for optimization, simultaneously achieving multi-scale predictions. YOLOv3-Tiny is a lightweight version of YOLOv3. It reduces the network's depth and width to achieve a more lightweight design. Fu et al. developed an improved YOLOv3-tiny model (DY3TNet) to detect kiwifruit in the orchard [132]. This method incorporated supplementary 3×3 and 1×1 convolutional kernels within the fifth and sixth convolution layers of the YOLOv3-tiny model to improve the model's ability to extract features while reducing computational complexity. Data augmentation techniques were used to increase the model's robustness. Experimental results indicated that the DY3TNet model achieved the highest average precision of 90.05% on the test images.

Ji et al. developed an apple recognition framework utilizing improved YOLOv4, wherein the lightweight EfficientNet-B0 network served as the feature extraction network integrated with the PANet (Path Aggregation Network) network for cross-layer feature fusion [133]. The same researchers further proposed a real-time apple detection methodology based on ShuffleNetV2-YOLOX, adapting YOLOX-Tiny as the baseline [134]. This implementation employed a ShuffleNetV2 backbone enhanced with Convolutional Block Attention Modules (CBAMs) and streamlined the network architecture to two feature extraction layers. Parallel advancements included Zhang et al.'s multi-class detection system for cherry tomatoes using refined YOLOv4-Tiny [135] and Hu et al.'s apple detection approach leveraging a modified YOLOX algorithm [136].

YOLOv5 transitions to the Pytorch framework to facilitate developer accessibility and extensibility. Yan et al. applied an improved YOLOv5s (the smallest version in the YOLOv5 series) to the detection of apple picking, which replaced the BottleneckCSP module with the BottleneckCSP-2 module, inserted an SE module to enhance feature extraction capabilities, and improved the feature map fusion method by effectively combining feature maps at different scales [137]. Kaukab et al. proposed a multimodal data fusion technology, i.e., the NBR-DF integrated with YOLOv5 for real-time apple fruit detection, which captured two different modalities of data (i.e., RGB and depth images) using a RealSense D455 stereo depth camera, and the NBR-DF method enhanced depth features by preprocessing the depth images, and then the depth images were combined with the RGB images and input into YOLOv5 for detection [138]. Lawal proposed a lightweight network for fruit instance segmentation, i.e., YOLOv5-LiNet, to strengthen fruit detection [139]. YOLOv5-LiNet employed a backbone comprising Stem, Shuffle_Block, ResNet, and SPPF modules and utilized PANet as the neck network and EIoU loss function to optimize detection performance. Tao et al. proposed STBNA-YOLOv5 for weed detection in rapeseed fields [140]. Their approach incorporated a Swin Transformer encoder block to augment feature extraction capabilities, integrated a BiFPN structure combined with a Normalization-based Attention Module (NAM) to facilitate efficient multi-scale feature utilization, and implemented an adaptive spatial fusion module to improve recognition sensitivity. Similarly, an enhanced YOLOv5-based methodology has been proposed for apple grading, specifically aiming

to mitigate limitations associated with low grading accuracy and suboptimal processing speed encountered during conventional grading procedures [141].

YOLOv7 improves multi-scale feature extraction, allowing it to handle objects of different sizes more effectively, especially enhancing its ability to detect small objects. It also adopts an adaptive anchor generation mechanism, enabling the model to better adapt to different datasets [142]. Gu et al. developed an improved citrus detection model based on the YOLOv7-tiny model, YOLO-DCA, which replaced the standard convolution in ELAN with depth-wise separable convolution to reduce the number of model parameters, combined the coordinate attention mechanism (CA) with standard convolution to form CAConv, and replaced the original detection head with a dynamic detection head to better detect targets of different scales [143].

YOLOv8 uses a more efficient backbone architecture, adopting a structure similar to CSPDarknet [144]. Compared to YOLOv5, it replaces all the C3 structures in both the backbone and neck with the C2f structure, which has richer gradient flow. The neck part incorporates PANet to handle multi-scale features, while in the head section, an anchor-free design is used to directly predict bounding box coordinates. Zhang et al. proposed a lightweight nectarine detection method based on YOLOv8n, i.e., YOLOv8n-CSD, to improve the accuracy of nectarine fruit detection in complex environments [145]. Ma et al. used an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments [146]. Ma et al. proposed an improved YOLOv8 model for instance segmentation of lotus seedpods in the lotus pond environment [147]. They integrated a convolutional block attention module (CBAM) within the neck network to extract and enhance discriminative feature extraction across maturity stages while maintaining computational efficiency and employed Wise-Intersection over Union (WIoU) as the regression loss function to minimize inference bias and improved the bounding box prediction accuracy. Parallel advancements included Xie et al.'s YOLOv8s-Seg-CBAM model for strawberry picking points localization [148].

Table 1. Comparison between major fruit detection approaches based on YOLO series.

Algorithms	Dataset	Feature	mAP/%	Speed (s/pic)	Reference
YOLOv3	RGB image (kiwifruit)	YOLOv3 uses Darknet-53 as the backbone network and extensively employs residual skip connections for optimization. But it has increased complexity, slower than easier version.	90.05	0.034	[132]
YOLOv4	RGB image (red apple)	YOLOv4 introduces CSPDarknet53 as the backbone and incorporates the SPP-block, while using PANet instead of FPN. But less user-friendly.	93.42	0.0158	[133]
YOLOv5	RGB-D image (red apple)	YOLOv5 moves to the Pytorch framework for developers to use and extend, making further improvements than YOLOV3. But less adaptable than YOLOV4.	96.4	0.01724	[138]
YOLOX	RGB-D Image (Fuji apple)	YOLOX employs anchor-free prediction and a decoupled head structure, introducing dynamic label assignment.	94.09	0.006	[136]

Table 1. Cont.

Algorithms	Dataset	Feature	mAP/%	Speed (s/pic)	Reference
YOLOv7	RGB image (Citrus)	YOLOv7 introduces the E-ELAN architecture, and the SPPCSPC module is improved. But it has higher complexity and, risk of overfitting.	96.98	0.0059	[143]
YOLOv8	RGB image (Over 300 apple varieties in multi-stage)	YOLOv8 uses a more efficient backbone, a structure similar to CSPDarknet. But it has case-specific optimization.	91.4	0.0254	[146]

(b) Fruit detection methods based on other algorithms.

SSD is a simple and efficient single-stage object detection framework, featuring multi-scale prediction, default boxes, and end-to-end training. It uses predefined anchor boxes with different aspect ratios and sizes and generates multiple bounding box predictions in a single forward pass, resulting in faster speed [149]. A method based on an improved SSD was proposed for detecting long jujubes in natural environments, which replaced the Peleenet network with VGG16 as the backbone network [150]. This method introduced a coordinate attention module and a global attention mechanism in the dense block to enhance target localization and recognition. Additionally, the Inceptionv2 module was replaced in the first three additional layers of the SSD structure, enhancing the model's ability to capture feature information. In related work, Agarwal et al. presented a deployment of the state-of-the-art SSD for detecting mango fruit on a tree canopy in orchard environments [151]. They utilized the deep learning model Darknet-19 as a feature extractor and SSD for object detection and localization.

DETR is a revolutionary object detection method that introduces the transformer architecture to handle the object detection task, simplifying many steps in traditional methods, especially by removing RPN and NMS. However, it has a high computational cost and longer training time, making it suitable for applications with sufficient computational resources [152]. It is also applied in fruit detection. Yang et al. proposed an improved DETR (named R2N-DETR) to detect multiple-sized peaches accurately in orchards with varying illumination and fruit occlusion, and Res2Net-50 was chosen as the backbone network to extract multi-scale convolutional features while maintaining a low resolution [153]. Huang et al. designed a lightweight transformer architecture derived from the RT-DETR model for robust pear fruit detection in a natural environment [154]. Ji et al. proposed a green apple detection method integrating a multidimensional feature extraction network model with a transformer module [155]. Their enhanced DETR architecture incorporated three key modifications: (1) utilization of ResNet18 as the primary feature extraction backbone, (2) strategic replacement of standard residual layers with deformable convolutions (DCNv2), and (3) implementation of scale-invariant attention mechanisms.

Zhu et al. proposed Olive-EfficientDet, a novel framework for multi-cultivar olive maturity classification in orchard environments [125]. The CBAM module was incorporated into the backbone network of EfficientDet, and an improved weighted Bi-FPN head network was proposed to focus on occluded and overlapping olive fruits. During the feature fusion stage, a weighting mechanism was introduced to assess the contribution of feature maps with different resolutions. The experimental results showed that the mean average precision (mAP) of fruit maturity detection reached 94.60%, 95.45%, 93.75%, and 96.05% for four olive varieties, and the mean detection time was 337 milliseconds per image.

Based on the aforementioned analysis, the advantages and disadvantages of the three types of fruit detection methods can be summarized in Table 2. Traditional image process-

ing methods have low computational requirements, fast processing speeds, and do not need a lot of training data. On the other side, they rely on manually designed feature extraction rules and exhibit poor robustness. Traditional machine methods have faster training speed, easier model interpretability, lower labeled data requirements, automatic feature learning, and stronger generalization ability. However, they are sensitive to the inputs of abnormal data and rely on manually designing and selecting appropriate feature extraction methods. Deep learning methods use end-to-end learning and an automatic feature extraction strategy, exhibit best robustness and adaptability in complex environments, and can adapt to multiple tasks and data types. In contrast, they require a large amount of labeled data for training and high computational resources and have poor model interpretability.

Table 2. Advantages and disadvantages of different fruit detection methods.

Methods	Algorithms	Advantages	Disadvantages
Traditional image processing	Color space transformation; Edge detection; Contour extraction; Threshold segmentation; Shape matching; Template matching.	Low computational requirements; Fast processing speeds; No need for a lot of training data.	Rely on manually designed feature extraction rules; Exhibit poor robustness.
Traditional machine learning	HOG-SVM; KNN; RF; Shape matching-SVM; DT; K-means.	Faster training speed; Easier model interpretability; Low labeled data requirements; Automatic feature learning; Stronger generalization ability.	Sensitive to the inputs of abnormal data; Rely on manually designing and selecting appropriate feature extraction methods; Limited robustness.
Deep learning	One-stage (YOLO, SSD, DETR, EfficientDet); Two-stage (Faster RCNN, Mask RCNN).	End-to-end learning; Automatic feature extraction; Exhibit best robustness and adaptability in complex environments; Adaptability to multiple tasks and data types.	Require a large amount of labeled data for training; High computational resource demands; Poor model interpretability.

Fruit detection methods have significantly improved the accuracy and robustness of fruit detection in complex backgrounds. However, the movement paths of picking robotic arms are often occupied by dense networks of branches. These branches not only frequently obstruct fruits but also pose the main collision risk during the robotic arm’s grasping actions. Ignoring the existence of branch obstacles can easily lead to picking failures, fruit damage, or even equipment damage. Therefore, the precise identification and location of key environmental obstacles, such as branches, constitute another core function indispensable to the fruit visual recognition system. Compared to fruits with regular shapes and distinct features, branches typically present complex topological structures that are slender, curved, and intertwined and are easily confused with the background (such as soil or sky) or other green vegetation. Therefore, the next section will analyze the current research status of visual detection methods for branch obstacles.

3.4. Tree Branch Detection Methods

The branches of fruit trees intertwine, and the distribution of fruits is random. For branch detection and segmentation, there are mainly two key functions: (1) obstacle avoidance and path planning, and (2) locating the picking point. Branch detection helps the robot recognize the location and shape of the branches, enabling it to plan a reasonable picking path, avoid collisions with branches, and improve the success rate of fruit picking.

By detecting the branches, the robot can more accurately locate the hanging points of the fruits, thereby determining the picking points for more precise harvesting operations, reducing mis-picks and missed picks, and improving harvesting accuracy. Current research on this issue can be classified into two main kinds: detection based on images and detection based on point clouds [156].

3.4.1. Tree Branch Detection Based on Images

The tree branches detection methods based on images identify the branches with their features from the picture. Chen et al. segmented the partially occluded apple trees based on three supervised learning models: U-Net, DeepLabv3, and Pix2Pix [157]. While DeepLabv3 demonstrated superior binary accuracy, mean IoU, and boundary F1 score, Pix2Pix (without discriminator) and U-Net achieved higher Occluded branch recall. Li et al. applied DeepLabv3 for semantic segmentation of RGB images into three categories: background, fruit, and twig [158]. To segment calyxes, branches, and wires for achieving higher-level picking strategies of kiwifruits, DeepLabV3+ was employed to segment the fruit calyxes, branches, and wires, complemented by a progressive probabilistic Hough transform (PPHT) method for discrete wire pixel reconstruction to infer spatial distribution [159]. Lin et al. [160] proposed a 3D reconstruction of guava branches using instance segmentation and geometry analysis. A tiny Mask RCNN was used to detect guava branches and then convert them to 3D point clouds, and then a cylindrical segment fitting method was proposed to reconstruct branches from the point clouds [160]. Yang et al. proposed an integrated system based on the Mask RCNN and a branch segment merging algorithm to simultaneously detect and measure citrus fruits and branches. The average precision of branch recognition was 96.27% [161].

Wan et al. presented a real-time detection framework combining CNN with image processing [162]. Their branch-CNN framework localized bare branches preliminarily, while HSV-based background segmentation refined regional boundaries. Field validation in pomegranate orchards yielded 90.7% precision, 89% recall, and 90% F1 score. A lightweight attention Ghost-HRNet (AGHRNet) incorporated deep learning to delineate tree branches from the complex orchard background [163]. This method introduced an attention-driven Ghost block module to reduce model complexity while enhancing segmentation accuracy. Comparative evaluations against contemporary benchmarks demonstrated AGHRNet's superior performance, achieving elevated segmentation metrics with reduced computational footprint on proprietary datasets. Complementarily, the feature intersection and fusion transformer (FIT-transformer) network was proposed to segment branch-background boundaries, facilitating the identification of safe robotic catch-and-shake zones. The architecture integrated two specialized components: a diverse feature aggregation (DFA) and an attention feature fusion module (AFFM), collectively engineered to strengthen feature discriminability and establish robust agricultural perception models [164].

Within natural, unstructured orchard environments, branch occlusion presents significant challenges. Existing reconstruction methodologies predominantly address planar representations with minimal occlusion, while contemporary 3D tree modeling approaches lack optimization for harvesting-specific constraints requiring computational efficiency and precise localization. To address this limitation, Kok et al. introduced a novel framework for reconstructing occluded 3D arboreal structures from planar RGB-D imagery [165]. The framework comprised three core components: branch segmentation using Unet++, branch reconstruction using Point2Skeleton, and branch recovery using a novel obscured branch recovery (OBR) algorithm. Experimental validation demonstrated the framework's efficacy in reconstructing spatial topology for both visible and occluded branches from monocular

RGB-D input. This approach exhibited significant potential for implementation in robotic harvesting applications.

Generally, image-based studies can achieve rather high accuracy, lightweight, and occlusion, which can provide useful information for selective fruit harvesting. However, most of the studies focused on 2D segmentation and were hard to extract 3D information.

3.4.2. Tree Branch Detection Based on Point Clouds

The detection methods based on point clouds detect and reconstruct the branches directly from 3D point cloud data collected by a stereo camera or LiDAR. Digumarti et al. proposed an automated technique for segmenting vegetation point clouds—captured with terrestrial laser scanners—into branch and leaf components by leveraging geometry-derived features computed directly from unstructured point data [166]. This approach achieved a mean classification accuracy of 91% across simulated datasets of three broadleaf tree species. Ma et al. proposed an automated branch detection framework for dormant jujube tree pruning, integrating 3D reconstruction with deep learning [167]. Their pipeline acquired high-fidelity point clouds using RGB-D cameras, employed the SPGNet architecture for trunk and branch segmentation, and implemented DBSCAN clustering for branch quantification. Validation on a proprietary dormant jujube dataset yielded an accuracy of 0.83 and an Intersection over Union (IoU) score of 0.75. While these end-to-end methodologies demonstrate efficacy in processing morphologically complex point clouds, they incur substantial computational latency and exhibit dependency on high-resolution laser-scanned data. Furthermore, point cloud segmentation demands significantly greater effort than RGB image segmentation in both custom dataset curation and annotation.

To solve this issue, Westling et al. proposed a graph-based method for analyzing orchard tree structure, which can work on low-quality data captured by handheld or mobile LiDAR [168]. The method leveraged fundamental geometric properties inherent to tree-like architectures—specifically, the connectivity of woody networks—which exhibit robustness against noise, fidelity, and occlusion. Zhang et al. proposed a novel framework for reconstructing occluded branches from RGB-D images [156]. Their technique extends established 2D branch sensing paradigms into three dimensions by fusing point clouds derived from planar segmentation masks and depth images, thereby exploiting multi-view information. The pipeline employed DeeplabV3+ and Pix2pix models for segmentation mask generation and utilized the Fast Global Registration (FGR) for multi-view point cloud alignment. The results demonstrated that the proposed framework significantly enhanced occluded branches reconstruction, achieving superior output completeness and computational efficiency. These advancements established their applicability within natural orchard environments.

In general, the methods based on point clouds require high-quality data and often consume a long processing time. The devices and methods are mostly studied in a lab environment; few works have considered their application in natural orchards.

4. Challenges and Future Trends

With the gradually accelerating process of urbanization, the shortage of the labor force in agricultural areas has become increasingly serious. The application of advanced technologies introduces Agriculture 4.0, which aims to replace labor with intelligent devices. The selective fruit harvesting robots are the trend for automatic picking of high-quality fruits. Visual detection is the eyes for selective harvesting robots. Extensive research has been performed for detecting the fruits and branches, yielding good performance like high accuracy, low computation requirement, etc. This section discusses the challenges that need

attention in the application of visual detection methods, followed by the future trends, as shown in Figure 12.

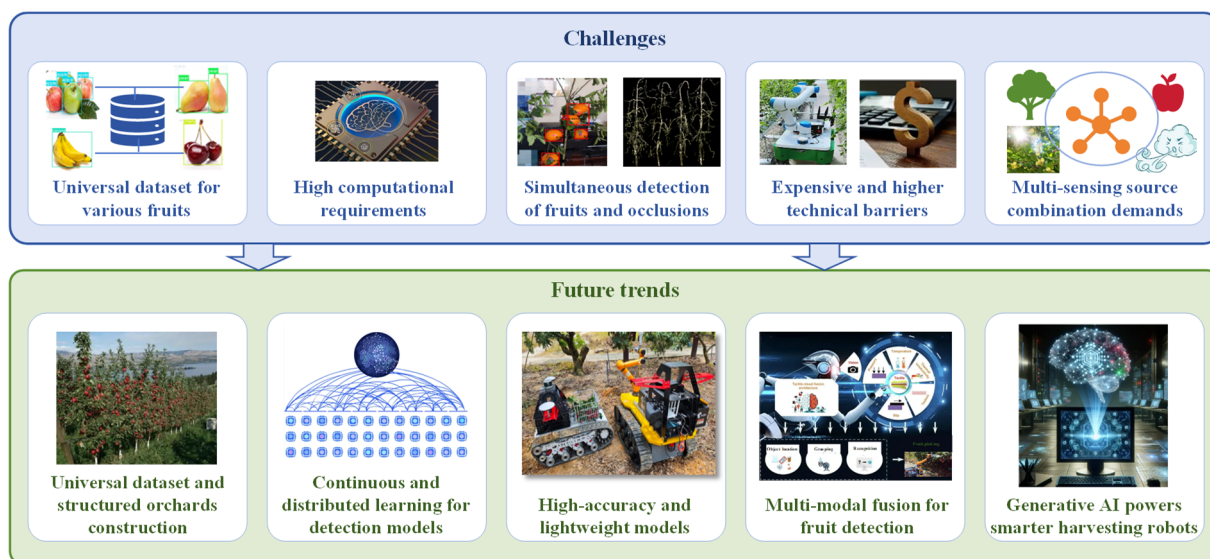


Figure 12. Challenges and future trends (some of the pictures are from Baidu).

4.1. Existing Challenges

(1) The dataset for a visual detection model is often constrained to one category of a certain fruit, which means that the learned model lacks universality. The dataset is the basis for learning a detection model. Since different fruits have different features, the structure and parameters of the visual detection models are often set or learned according to the specific kind of fruit. However, the fruits vary largely, e.g., the apple planting environment across various regions, seasons, and types, while the apple detection models can only work on a specific kind of apple. Consequently, the scalability and universality of the visual system should be further studied.

(2) The complexity of detection models often hinders their deployment on embedded computation devices with limited capacity. In natural orchards, fruits tend to overlap with branches and leaves, and the lighting conditions are uncertain. To achieve the target of high detection accuracy, the models often need a long time to learn and run; the real-time capabilities of the vision system are limited, leading to delays from detection to harvesting. Problems such as misrecognition, missed recognition, and inaccurate positioning often happen, especially when the fruit area is incomplete due to being blocked by branches and leaves, which may result in the failure picking of the end effector.

(3) Occlusion is one of the most typical problems of robotic harvesting since it significantly affects the detection of the targeted fruits and constrains the occlusion of the picking path of the end effectors. In the existing studies, the fruits and tree branches are often detected by different models; the entire detection for the whole environment for fruit growth is hard to achieve by one model. The reconstruction of tree branches is seen as one of the key issues for robotic fruit detection, but its large computational requirement is hard to conquer.

(4) The visual detection system is hard to apply and operate by the farmers on the current version. Since the key devices of the selective harvesting robots (e.g., camera, embedded devices, and robot arm) are often expensive, the selective harvesting robots are often hard for farmers to afford. In addition, the model learning and debugging often need a background in computer knowledge and artificial intelligence, and the cost of operation and maintenance of the system is also high. For example, the stable power, uninterrupted

Internet connection, and regular equipment calibration are often hard to maintain in the field environment.

(5) In the natural orchards, amounts of multi-source heterogeneous data are collected by the smart sensors, including fruit and tree shape, light change, wind strength, etc. The single visual sensor can hardly keep working stably, and the detection based on sole visual information may provide low-quality results in recognition and localization for fruit picking. The detection methods combining multi-source data have the potential to obtain accurate and stable support for selective fruit harvesting robots.

4.2. Future Trends

To address the persistent challenges confronting the visual system in current selective fruit harvesting robots, further studies should prioritize advancements in universality, high efficiency, low cost, and easy usability. Therefore, the future trends can evolve around the following aspects.

(1) In order to improve the scalability of the detection system, the universality of the raw datasets should be enhanced, for example, by expanding the fruit datasets of different regions, different growth stages, and different varieties to make them applicable to a wider range of application scenarios. At the same time, structured orchards can be trimmed to minimize or eliminate overlapping fruits, including appropriate thinning, leaf management, etc. The trim procedures should collaborate with horticultural scientists and agricultural engineers to reduce the fruit detection complexity without losing fruit yield.

(2) To improve the sustainability and effectiveness of the visual detection models, continuous learning and discrete learning mechanisms need to be studied. The models should adjust their parameters to meet the changes in light variations, differences in fruit growth, and occlusion issues. Through self-supervised learning and incremental fine-tuning, the models can be updated online to adapt to a new fruit growth environment. In addition, based on a distributed collaborative optimization (e.g., federated learning) framework, multiple harvesting robots can share local fruit detection model parameters to integrate diverse scene data while enhancing generalization.

(3) Further refining algorithms is still a trend to improve the performance of visual fruit detection systems in natural environments. The accuracy and the lightweight are the two main optimization objectives. For the accuracy aspect, more new published deep learning structures and modules in the computer algorithm field (e.g., visual prompting and large visual models) can be introduced in the fruit detection area. Similarly, more model compression (e.g., light convolution) and edge deployment optimization (e.g., neural architecture search) technologies can be applied for the lightweight target. In addition, the newer lightweight RGB-D cameras suitable for embedded robotic platforms (e.g., Luxonis OAK-D) can also be studied.

(4) The evolution of fruit detection robots is taking full-modal real-time closed-loop (deep fusion of mechanical, spectral, and environmental data), edge-cloud co-evolution (lightweight model and continuous learning), and high-fidelity simulation (physics engine and domain adaptation) as the core paths. Multimodal data (including tactile, image, text, language interaction signals, wind speed, etc.) are widely sensed and can be integrated for fruit detection. The visual fruit detection system should evolve to follow the main direction with the multimodal data fusion technologies, such as dynamic uncertainty-aware fusion (e.g., DUA-Net), Mamba-guided multimodal reconstruction (MGMR-Net), etc.

(5) Generative AI significantly empowers smarter fruit harvesting robots by accelerating robot design, optimizing detection capabilities, and enabling adaptive autonomy in dynamic orchards. Multi-model large language models (LLMs) can help robots understand the orchard environment and provide humanity and assistance for making more reasonable

picking tasks. Vision language models (VLMs) can construct a bridge between visual detection and language understanding, empowering the harvesting robots to understand the orchards visually and respond with interpretive languages.

5. Conclusions

With the development of advanced technologies, the application of selective fruit harvesting robots is expected as an important direction for future agriculture. A visual fruit detection system is essential to provide useful fruit information for the path optimization of the arm and the picking activity of the end effector. The reliability of the visual detection system should take into account the working scene, appropriate vision sensor selection, and visual detection algorithms.

This paper reviewed the recent progress and development of visual detection methods for selective fruit harvesting robots. The cameras, the fruit detection methods, and the tree branch detection methods are emphatically introduced. (1) The cameras primarily fall into three distinct categories based on their operating principles: binocular stereo cameras, structured light cameras, and ToF cameras. Their principles, recent applications, advantages, and disadvantages are analyzed. (2) The fruit detection methods are categorized into two paradigms: traditional methods and deep learning-based methods. The former relies on manually designed feature extraction algorithms and is mainly suitable for fruits whose growth environments are relatively simple and predictable, and they often have lower computational requirements and faster processing speeds. The latter applies the deep learning methods for detecting fruits, which can directly learn the final fruit detection task objective from raw images, and can be classified into two main directions, i.e., two-stage models and one-stage models. They can obtain high detection accuracy of fruits in complex unstructured orchards. (3) Tree branch detection is also important during fruit harvesting, which mainly relies on images or point clouds. The former mainly focused on 2D segmentation and can achieve 3D information by further integrating with a depth image. The latter can directly obtain the 3D information but with a high computation load and cost. Finally, the potential challenges and future trends of the visual detection system of selective fruit harvesting robots are analyzed to provide readers with a full understanding of the latest progress in this field.

Author Contributions: Conceptualization, W.W. and M.Z.; methodology, C.L. and M.Z.; software, C.L. and Y.P.; validation, C.L. and X.Z.; formal analysis, W.W. and C.L.; investigation, Y.X.; resources, J.G.; data curation, J.G.; writing—original draft preparation, W.W. and Y.X.; writing—review and editing, W.W. and J.G.; visualization, C.L. and Y.X.; supervision, X.Z. and M.Z.; project administration, J.G. and X.Z.; funding acquisition, W.W. and Y.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (52105516), the special scientific research project of the School of Emergency Management, Jiangsu University (KY-A-03), and the Yongjia Agriculture and Rural Bureau Science and Technology Special Project (No. 2024YJ004).

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Y.; Ma, X.; Shu, L.; Hancke, G.P.; Abu-Mahfouz, A.M. From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4322–4334. [[CrossRef](#)]

2. Lakhiar, I.A.; Yan, H.; Zhang, C.; Wang, G.; He, B.; Hao, B.; Han, Y.; Wang, B.; Bao, R.; Syed, T.N.; et al. A Review of Precision Irrigation Water-Saving Technology under Changing Climate for Enhancing Water Use Efficiency, Crop Yield, and Environmental Footprints. *Agriculture* **2024**, *14*, 1141. [\[CrossRef\]](#)
3. Ouafiq, E.M.; Saadane, R.; Chehri, A.; Jeon, S. AI-Based Modeling and Data-Driven Evaluation for Smart Farming-Oriented Big Data Architecture Using IoT with Energy Harvesting Capabilities. *Sustain. Energy Technol. Assess* **2022**, *52*, 102093. [\[CrossRef\]](#)
4. Wang, G.; Zhang, C.; Liu, S.; Zhao, Y.; Zhang, Y.; Wang, L. Multi-Robot Collaborative Manufacturing Driven by Digital Twins: Advancements, Challenges, and Future Directions. *J. Manuf. Syst.* **2025**, *82*, 333–361. [\[CrossRef\]](#)
5. Wang, W.; Yang, S.; Zhang, X.; Xia, X. Research on the Smart Broad Bean Harvesting System and the Self-Adaptive Control Method Based on CPS Technologies. *Agronomy* **2024**, *14*, 1405. [\[CrossRef\]](#)
6. Ma, S.; Ding, W.; Liu, Y.; Ren, S.; Yang, H. Digital Twin and Big Data-Driven Sustainable Smart Manufacturing Based on Information Management Systems for Energy-Intensive Industries. *Appl. Energy* **2022**, *326*, 119986. [\[CrossRef\]](#)
7. Wang, W.; Shan, Y.; Xi, Y.; Xia, Z.; Xu, G.; Zhang, X. A Predictive Production-Logistics Cooperation Method for Service-Oriented Smart Discrete Manufacturing System. *J. Eng. Des.* **2025**. [\[CrossRef\]](#)
8. Xia, Z.; Zhao, Y.; Gu, J.; Wang, W.; Zhang, W.; Huang, Z. FC-DETR: High-Precision End-to-End Surface Defect Detector Based on Foreground Supervision and Cascade Refined Hybrid Matching. *Expert Syst. Appl.* **2025**, *266*, 126142. [\[CrossRef\]](#)
9. Zhu, X.; Chikangaise, P.; Shi, W.; Chen, W.-H.; Yuan, S. Review of Intelligent Sprinkler Irrigation Technologies for Remote Autonomous System. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 23–30. [\[CrossRef\]](#)
10. Darko, R.O.; Yuan, S.; Hong, L.; Liu, J.; Yan, H. Irrigation, a Productive Tool for Food Security—A Review. *Acta. Agric. Scand. B Soil Plant Sci.* **2016**, *66*, 191–206. [\[CrossRef\]](#)
11. Hou, P.; Yuan, W.; Li, G.; Petropoulos, E.; Xue, L.; Feng, Y.; Xue, L.; Yang, L.; Ding, Y. Deep Fertilization with Controlled-release Fertilizer for Higher Cereal Yield and N Utilization in Paddies: The Optimal Fertilization Depth. *Agron. J.* **2021**, *113*, 5027–5039. [\[CrossRef\]](#)
12. Sorour, S.E.; Alsayyari, M.; Alqahtani, N.; Aldosery, K.; Altaweel, A.; Alzhrani, S. An Intelligent Management System and Advanced Analytics for Boosting Date Production. *Sustainability* **2025**, *17*, 5636. [\[CrossRef\]](#)
13. Wu, M.; Liu, S.; Li, Z.; Ou, M.; Dai, S.; Dong, X.; Wang, X.; Jiang, L.; Jia, W. A Review of Intelligent Orchard Sprayer Technologies: Perception, Control, and System Integration. *Horticulturae* **2025**, *11*, 668. [\[CrossRef\]](#)
14. Subeesh, A.; Prakash Kumar, S.; Kumar Chakraborty, S.; Upendar, K.; Singh Chandel, N.; Jat, D.; Dubey, K.; Modi, R.U.; Mazhar Khan, M. UAV Imagery Coupled Deep Learning Approach for the Development of an Adaptive In-House Web-Based Application for Yield Estimation in Citrus Orchard. *Measurement* **2024**, *234*, 114786. [\[CrossRef\]](#)
15. Liu, J.; Liang, J.; Zhao, S.; Jiang, Y.; Wang, J.; Jin, Y. Design of a Virtual Multi-Interaction Operation System for Hand-Eye Coordination of Grape Harvesting Robots. *Agronomy* **2023**, *13*, 829. [\[CrossRef\]](#)
16. Xu, Z.; Liu, J.; Wang, J.; Cai, L.; Jin, Y.; Zhao, S.; Xie, B. Realtime Picking Point Decision Algorithm of Trellis Grape for High-Speed Robotic Cut-and-Catch Harvesting. *Agronomy* **2023**, *13*, 1618. [\[CrossRef\]](#)
17. Ma, J.; Li, M.; Fan, W.; Liu, J. State-of-the-Art Techniques for Fruit Maturity Detection. *Agronomy* **2024**, *14*, 2783. [\[CrossRef\]](#)
18. Herman, R.A.; Ayepa, E.; Fometu, S.S.; Shittu, S.; Davids, J.S.; Wang, J. Mulberry Fruit Post-Harvest Management: Techniques, Composition and Influence on Quality Traits—A Review. *Food Control* **2022**, *140*, 109126. [\[CrossRef\]](#)
19. Ji, W.; Qian, Z.; Xu, B.; Tang, W.; Li, J.; Zhao, D. Grasping Damage Analysis of Apple by End-effector in Harvesting Robot. *J. Food Process Eng.* **2017**, *40*, e12589. [\[CrossRef\]](#)
20. Au, W.; Zhou, H.; Liu, T.; Kok, E.; Wang, X.; Wang, M.; Chen, C. The Monash Apple Retrieving System: A Review on System Intelligence and Apple Harvesting Performance. *Comput. Electron. Agric.* **2023**, *213*, 108164. [\[CrossRef\]](#)
21. Wang, J.; Zhang, Y.; Gu, R. Research Status and Prospects on Plant Canopy Structure Measurement Using Visual Sensors Based on Three-Dimensional Reconstruction. *Agriculture* **2020**, *10*, 462. [\[CrossRef\]](#)
22. Zhang, X.; Wang, H.; Dong, H. A Survey of Deep Learning-Driven 3D Object Detection: Sensor Modalities, Technical Architectures, and Applications. *Sensors* **2025**, *25*, 3668. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Chen, K.; Li, T.; Yan, T.; Xie, F.; Feng, Q.; Zhu, Q.; Zhao, C. A Soft Gripper Design for Apple Harvesting with Force Feedback and Fruit Slip Detection. *Agriculture* **2022**, *12*, 1802. [\[CrossRef\]](#)
24. Lu, E.; Ma, Z.; Li, Y.; Xu, L.; Tang, Z. Adaptive Backstepping Control of Tracked Robot Running Trajectory Based on Real-Time Slip Parameter Estimation. *Int. J. Agric. Biol. Eng.* **2020**, *13*, 178–187. [\[CrossRef\]](#)
25. Magalhães, S.A.; Moreira, A.P.; dos Santos, F.N.; Dias, J. Active Perception Fruit Harvesting Robots—A Systematic Review. *J. Intell. Robot. Syst. Theory Appl.* **2022**, *105*, 14. [\[CrossRef\]](#)
26. Luo, Y.; Li, J.; Yao, B.; Luo, Q.; Zhu, Z.; Wu, W. Research Progress and Development Trend of Bionic Harvesting Technology. *Comput. Electron. Agric.* **2024**, *222*, 109013. [\[CrossRef\]](#)
27. Qureshi, W.S.; Payne, A.; Walsh, K.B.; Linker, R.; Cohen, O.; Dailey, M.N. Machine Vision for Counting Fruit on Mango Tree Canopies. *Precis. Agric.* **2017**, *18*, 224–244. [\[CrossRef\]](#)

28. Jia, W.; Zhang, Y.; Lian, J.; Zheng, Y.; Zhao, D.; Li, C. Apple Harvesting Robot under Information Technology: A Review. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 1–16. [\[CrossRef\]](#)
29. Khan, Z.; Shen, Y.; Liu, H. Object Detection in Agriculture: A Comprehensive Review of Methods, Applications, Challenges, and Future Directions. *Agriculture* **2025**, *15*, 1351. [\[CrossRef\]](#)
30. Jin, Y.; Liu, J.; Xu, Z.; Yuan, S.; Li, P.; Wang, J. Development Status and Trend of Agricultural Robot Technology. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 1–19. [\[CrossRef\]](#)
31. Wang, H.; Gu, J.; Wang, M. A Review on the Application of Computer Vision and Machine Learning in the Tea Industry. *Front. Sustain. Food Syst.* **2023**, *7*, 1172543. [\[CrossRef\]](#)
32. Bai, Y.; Zhang, B.; Xu, N.; Zhou, J.; Shi, J.; Diao, Z. Vision-Based Navigation and Guidance for Agricultural Autonomous Vehicles and Robots: A Review. *Comput. Electron. Agric.* **2023**, *205*, 107584. [\[CrossRef\]](#)
33. Xiao, F.; Wang, H.; Li, Y.; Cao, Y.; Lv, X.; Xu, G. Object Detection and Recognition Techniques Based on Digital Image Processing and Traditional Machine Learning for Fruit and Vegetable Harvesting Robots: An Overview and Review. *Agronomy* **2023**, *13*, 639. [\[CrossRef\]](#)
34. Hua, W.; Zhang, Z.; Zhang, W.; Liu, X.; Hu, C.; He, Y.; Mhamed, M.; Li, X.; Dong, H.; Saha, C.K.; et al. Key Technologies in Apple Harvesting Robot for Standardized Orchards: A Comprehensive Review of Innovations, Challenges, and Future Directions. *Comput. Electron. Agric.* **2025**, *235*, 110343. [\[CrossRef\]](#)
35. Wu, H.; Wang, X.; Chen, X.; Zhang, Y.; Zhang, Y. Review on Key Technologies for Autonomous Navigation in Field Agricultural Machinery. *Agriculture* **2025**, *15*, 1297. [\[CrossRef\]](#)
36. Ma, S.; Ding, W.; Liu, Y.; Zhang, Y.; Ren, S.; Kong, X.; Leng, J. Industry 4.0 and Cleaner Production: A Comprehensive Review of Sustainable and Intelligent Manufacturing for Energy-Intensive Manufacturing Industries. *J. Clean. Prod.* **2024**, *467*, 142879. [\[CrossRef\]](#)
37. Zhou, H.; Wang, X.; Au, W.; Kang, H.; Chen, C. Intelligent Robots for Fruit Harvesting: Recent Developments and Future Challenges. *Precis. Agric.* **2022**, *23*, 1856–1907. [\[CrossRef\]](#)
38. Zhou, J.; He, L.; Karkee, M.; Zhang, Q. Analysis of Shaking-Induced Cherry Fruit Motion and Damage. *Biosyst. Eng.* **2016**, *144*, 105–114. [\[CrossRef\]](#)
39. Sola-Guirado, R.R.; Castro-Garcia, S.; Blanco-Roldán, G.L.; Gil-Ribes, J.A.; González-Sánchez, E.J. Performance Evaluation of Lateral Canopy Shakers with Catch Frame for Continuous Harvesting of Oranges for Juice Industry. *Int. J. Agric. Biol. Eng.* **2020**, *13*, 88–93. [\[CrossRef\]](#)
40. Wang, W.; Lu, H.; Zhang, S.; Yang, Z. Damage Caused by Multiple Impacts of Litchi Fruits during Vibration Harvesting. *Comput. Electron. Agric.* **2019**, *162*, 732–738. [\[CrossRef\]](#)
41. Han, L.; Mao, H.; Kumi, F.; Hu, J. Development of a Multi-Task Robotic Transplanting Workcell for Greenhouse Seedlings. *Appl. Eng. Agric.* **2018**, *34*, 335–342. [\[CrossRef\]](#)
42. Levin, M.; Degani, A. Design of a Task-Based Modular Re-Configurable Agricultural Robot. *IFAC-PapersOnLine* **2016**, *49*, 184–189. [\[CrossRef\]](#)
43. Barnett, J.; Duke, M.; Au, C.K.; Lim, S.H. Work Distribution of Multiple Cartesian Robot Arms for Kiwifruit Harvesting. *Comput. Electron. Agric.* **2020**, *169*, 105202. [\[CrossRef\]](#)
44. Zhang, K.; Lammers, K.; Chu, P.; Li, Z.; Lu, R. System Design and Control of an Apple Harvesting Robot. *Mechatronics* **2021**, *79*, 102644. [\[CrossRef\]](#)
45. Japar, A.F.; Ramli, H.R.; Norsahperi, N.M.H.; Hasan, W.Z.W. Oil Palm Loose Fruit Detection Using YOLOv4 for an Autonomous Mobile Robot Collector. *IEEE Access* **2024**, *12*, 138582–138593. [\[CrossRef\]](#)
46. Xiao, X.; Wang, Y.; Zhou, B.; Jiang, Y. Flexible Hand Claw Picking Method for Citrus-Picking Robot Based on Target Fruit Recognition. *Agriculture* **2024**, *14*, 1227. [\[CrossRef\]](#)
47. Khoshrangbaf, M.; Akram, V.K.; Challenger, M.; Dagdeviren, O. An Experimental Evaluation of Indoor Localization in Autonomous Mobile Robots. *Sensors* **2025**, *25*, 2209. [\[CrossRef\]](#)
48. Xie, F.; Guo, Z.; Li, T.; Feng, Q.; Zhao, C. Dynamic Task Planning for Multi-Arm Harvesting Robots Under Multiple Constraints Using Deep Reinforcement Learning. *Horticulturae* **2025**, *11*, 88. [\[CrossRef\]](#)
49. Yu, Y.; Xie, H.; Zhang, K.; Wang, Y.; Li, Y.; Zhou, J.; Xu, L. Design, Development, Integration, and Field Evaluation of a Ridge-Planting Strawberry Harvesting Robot. *Agriculture* **2024**, *14*, 2126. [\[CrossRef\]](#)
50. Liu, H.; Yan, S.; Shen, Y.; Li, C.; Zhang, Y.; Hussain, F. Model Predictive Control System Based on Direct Yaw Moment Control for 4wid Self-Steering Agriculture Vehicle. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 175–181. [\[CrossRef\]](#)
51. Huang, M.; Jiang, X.; He, L.; Choi, D.; Pecchia, J.; Li, Y. Development of a Robotic Harvesting Mechanism for Button Mushrooms. *Trans. ASABE* **2021**, *64*, 565–575. [\[CrossRef\]](#)
52. Ji, W.; He, G.; Xu, B.; Zhang, H.; Yu, X. A New Picking Pattern of a Flexible Three-Fingered End-Effector for Apple Harvesting Robot. *Agriculture* **2024**, *14*, 102. [\[CrossRef\]](#)

53. Pi, J.; Liu, J.; Zhou, K.; Qian, M. An Octopus-Inspired Bionic Flexible Gripper for Apple Grasping. *Agriculture* **2021**, *11*, 1014. [\[CrossRef\]](#)
54. Zhou, K.; Xia, L.; Liu, J.; Qian, M.; Pi, J. Design of a Flexible End-Effector Based on Characteristics of Tomatoes. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 13–24. [\[CrossRef\]](#)
55. Liu, J.; Yuan, Y.; Gao, Y.; Tang, S.; Li, Z. Virtual Model of Grip-and-Cut Picking for Simulation of Vibration and Falling of Grape Clusters. *Trans. ASABE* **2019**, *62*, 603–614. [\[CrossRef\]](#)
56. Faheem, M.; Liu, J.; Chang, G.; Ahmad, I.; Peng, Y. Hanging Force Analysis for Realizing Low Vibration of Grape Clusters during Speedy Robotic Post-Harvest Handling. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 62–71. [\[CrossRef\]](#)
57. Yang, N.; Chang, K.; Dong, S.; Tang, J.; Wang, A.; Huang, R.; Jia, Y. Rapid Image Detection and Recognition of Rice False Smut Based on Mobile Smart Devices with Anti-Light Features from Cloud Database. *Biosyst. Eng.* **2022**, *218*, 229–244. [\[CrossRef\]](#)
58. Jia, W.; Zheng, Y.; Zhao, D.; Yin, X.; Liu, X.; Du, R. Preprocessing Method of Night Vision Image Application in Apple Harvesting Robot. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 158–163. [\[CrossRef\]](#)
59. Huang, X.; Pan, S.; Sun, Z.; Ye, W.; Aheto, J.H. Evaluating Quality of Tomato during Storage Using Fusion Information of Computer Vision and Electronic Nose. *J. Food Process Eng.* **2018**, *41*, e12832. [\[CrossRef\]](#)
60. Huang, X.; Lv, R.; Wang, S.; Aheto, J.H.; Dai, C. Integration of Computer Vision and Colorimetric Sensor Array for Nondestructive Detection of Mango Quality. *J. Food Process Eng.* **2018**, *41*, e12873. [\[CrossRef\]](#)
61. Zhang, Z.; Lu, Y.; Yang, M.; Wang, G.; Zhao, Y.; Hu, Y. Optimal Training Strategy for High-Performance Detection Model of Multi-Cultivar Tea Shoots Based on Deep Learning Methods. *Sci. Hortic.* **2024**, *328*, 112949. [\[CrossRef\]](#)
62. Lu, P.; Zheng, W.; Lv, X.; Xu, J.; Zhang, S.; Li, Y.; Zhangzhong, L. An Extended Method Based on the Geometric Position of Salient Image Features: Solving the Dataset Imbalance Problem in Greenhouse Tomato Growing Scenarios. *Agriculture* **2024**, *14*, 1893. [\[CrossRef\]](#)
63. Altaheri, H.; Alsulaiman, M.; Muhammad, G.; Amin, S.U.; Bencherif, M.; Mekhtiche, M. Date Fruit Dataset for Intelligent Harvesting. *Data Brief* **2019**, *26*, 104514. [\[CrossRef\]](#)
64. James, J.A.; Manching, H.K.; Mattia, M.R.; Bowman, K.D.; Hulse-Kemp, A.M.; Beksi, W.J. CitDet: A Benchmark Dataset for Citrus Fruit Detection. *IEEE Robot. Autom. Lett.* **2024**, *9*, 10788–10795. [\[CrossRef\]](#)
65. Azizi, A.; Zhang, Z.; Hua, W.; Li, M.; Igathinathane, C.; Yang, L.; Ampatzidis, Y.; Ghasemi-Varnamkhashi, M.; Radi; Zhang, M.; et al. Image Processing and Artificial Intelligence for Apple Detection and Localization: A Comprehensive Review. *Comput. Sci. Rev.* **2024**, *54*, 100690. [\[CrossRef\]](#)
66. Zhu, J.; Jiang, X.; Rong, Y.; Wei, W.; Wu, S.; Jiao, T.; Chen, Q. Label-free detection of trace level zearalenone in corn oil by surface-enhanced Raman spectroscopy (SERS) coupled with deep learning models. *Food Chem.* **2023**, *414*, 135705. [\[CrossRef\]](#)
67. Wang, C.; Zou, X.; Tang, Y.; Luo, L.; Feng, W. Localisation of Litchi in an Unstructured Environment Using Binocular Stereo Vision. *Biosyst. Eng.* **2016**, *145*, 39–51. [\[CrossRef\]](#)
68. Wang, J.; Gao, Z.; Zhang, Y.; Zhou, J.; Wu, J.; Li, P. Real-Time Detection and Location of Potted Flowers Based on a ZED Camera and a YOLO V4-Tiny Deep Learning Algorithm. *Horticulturae* **2021**, *8*, 21. [\[CrossRef\]](#)
69. Tang, Y.; Zhou, H.; Wang, H.; Zhang, Y. Fruit Detection and Positioning Technology for a Camellia Oleifera C. Abel Orchard Based on Improved YOLOv4-Tiny Model and Binocular Stereo Vision. *Expert Syst. Appl.* **2023**, *211*, 118573. [\[CrossRef\]](#)
70. Li, T.; Fang, W.; Zhao, G.; Gao, F.; Wu, Z.; Li, R.; Fu, L.; Dhupia, J. An Improved Binocular Localization Method for Apple Based on Fruit Detection Using Deep Learning. *Inf. Process. Agric.* **2023**, *10*, 276–287. [\[CrossRef\]](#)
71. Zhang, L.; Hao, Q.; Mao, Y.; Su, J.; Cao, J. Beyond Trade-Off: An Optimized Binocular Stereo Vision Based Depth Estimation Algorithm for Designing Harvesting Robot in Orchards. *Agriculture* **2023**, *13*, 1117. [\[CrossRef\]](#)
72. Zhang, H.; Zhang, J.; Zhang, Y.; Wei, J.; Zhan, B.; Liu, X.; Luo, W. Structured-Illumination Reflectance Imaging Combined with Deep Learning for Detecting Early Decayed Oranges. *Postharvest Biol. Technol.* **2024**, *217*, 113121. [\[CrossRef\]](#)
73. Zhang, Q.; Su, W.H. Real-Time Recognition and Localization of Apples for Robotic Picking Based on Structural Light and Deep Learning. *Smart Cities* **2023**, *6*, 3393–3410. [\[CrossRef\]](#)
74. Li, T.; Feng, Q.; Qiu, Q.; Xie, F.; Zhao, C. Occluded Apple Fruit Detection and Localization with a Frustum-Based Point-Cloud-Processing Approach for Robotic Harvesting. *Remote Sens.* **2022**, *14*, 482. [\[CrossRef\]](#)
75. Yang, N.; Yuan, M.; Wang, P.; Zhang, R.; Sun, J.; Mao, H. Tea diseases detection based on fast infrared thermal image processing technology. *J. Sci. Food Agric.* **2019**, *99*, 3459–3466. [\[CrossRef\]](#)
76. Sun, M.; Xu, L.; Luo, R.; Lu, Y.; Jia, W. Fast Location and Recognition of Green Apple Based on RGB-D Image. *Front. Plant Sci.* **2022**, *13*, 864458. [\[CrossRef\]](#)
77. Legg, M.; Parr, B.; Pascual, G.; Alam, F. Grape Maturity Estimation Using Time-of-Flight and LiDAR Depth Cameras. *Sensors* **2024**, *24*, 5109. [\[CrossRef\]](#)
78. Peebles, M.; Lim, S.H.; Duke, M.; McGuinness, B.; Au, C.K. Localization of Asparagus Spears Using Time-of-Flight Imaging for Robotic Harvesting. *Ind. Robot.* **2024**, *51*, 595–606. [\[CrossRef\]](#)

79. Elbeltagi, A.; Srivastava, A.; Deng, J.; Li, Z.; Raza, A.; Khadke, L.; Yu, Z.; El-Rawy, M. Forecasting vapor pressure deficit for agricultural water management using machine learning in semi-arid environments. *Agric. Water Manag.* **2023**, *283*, 108302. [[CrossRef](#)]
80. Goel, N.; Sehgal, P. Fuzzy Classification of Pre-Harvest Tomatoes for Ripeness Estimation—An Approach Based on Automatic Rule Learning Using Decision Tree. *Appl. Soft. Comput.* **2015**, *36*, 45–56. [[CrossRef](#)]
81. Yu, X.; Fan, Z.; Wang, X.; Wan, H.; Wang, P.; Zeng, X.; Jia, F. A Lab-Customized Autonomous Humanoid Apple Harvesting Robot. *Comput. Electr. Eng.* **2021**, *96*, 107459. [[CrossRef](#)]
82. Ratprakhon, K.; Neubauer, W.; Riehn, K.; Fritsche, J.; Rohn, S. Developing an Automatic Color Determination Procedure for the Quality Assessment of Mangos (*Mangifera Indica*) Using a CCD Camera and Color Standards. *Foods* **2020**, *9*, 1709. [[CrossRef](#)]
83. Tan, K.; Lee, W.S.; Gan, H.; Wang, S. Recognising Blueberry Fruit of Different Maturity Using Histogram Oriented Gradients and Colour Features in Outdoor Scenes. *Biosyst. Eng.* **2018**, *176*, 59–72. [[CrossRef](#)]
84. Li, L.; Xie, S.; Zhu, F.; Ning, J.; Chen, Q.; Zhang, Z. Colorimetric Sensor Array-Based Artificial Olfactory System for Sensing Chinese Green Tea's Quality: A Method of Fabrication. *Int. J. Food Prop.* **2017**, *20*, 1762–1773. [[CrossRef](#)]
85. Ji, W.; Qian, Z.; Xu, B.; Zhao, D. A Nighttime Image Enhancement Method Based on Retinex and Guided Filter for Object Recognition of Apple Harvesting Robot. *Int. J. Adv. Robot. Syst.* **2018**, *15*, 1–12. [[CrossRef](#)]
86. Lin, G.; Tang, Y.; Zou, X.; Cheng, J.; Xiong, J. Fruit Detection in Natural Environment Using Partial Shape Matching and Probabilistic Hough Transform. *Precis. Agric.* **2020**, *21*, 160–177. [[CrossRef](#)]
87. Lv, J.; Ni, H.; Wang, Q.; Yang, B.; Xu, L. A Segmentation Method of Red Apple Image. *Sci. Hortic.* **2019**, *256*, 108615. [[CrossRef](#)]
88. Oo, L.M.; Aung, N.Z. A Simple and Efficient Method for Automatic Strawberry Shape and Size Estimation and Classification. *Biosyst. Eng.* **2018**, *170*, 96–107. [[CrossRef](#)]
89. Lu, J.; Lee, W.S.; Gan, H.; Hu, X. Immature Citrus Fruit Detection Based on Local Binary Pattern Feature and Hierarchical Contour Analysis. *Biosyst. Eng.* **2018**, *171*, 78–90. [[CrossRef](#)]
90. Liu, X.; Jia, W.; Ruan, C.; Zhao, D.; Gu, Y.; Chen, W. The Recognition of Apple Fruits in Plastic Bags Based on Block Classification. *Precis. Agric.* **2018**, *19*, 735–749. [[CrossRef](#)]
91. Wu, J.; Zhang, B.; Zhou, J.; Xiong, Y.; Gu, B.; Yang, X. Automatic Recognition of Ripening Tomatoes by Combining Multi-Feature Fusion with a Bi-Layer Classification Strategy for Harvesting Robots. *Sensors* **2019**, *19*, 612. [[CrossRef](#)] [[PubMed](#)]
92. Liu, X.; Zhao, D.; Jia, W.; Ji, W.; Sun, Y. A Detection Method for Apple Fruits Based on Color and Shape Features. *IEEE Access* **2019**, *7*, 67923–67933. [[CrossRef](#)]
93. Mustaffa, M.R.; Xin Yi, N.; Abdullah, L.N.; Nasharuddin, N.A. Durian Recognition Based on Multiple Features and Linear Discriminant Analysis. *Malays. J. Comput. Sci.* **2018**, *28*, 57–72. [[CrossRef](#)]
94. Liu, T.-H.; Ehsani, R.; Toudeshki, A.; Zou, X.-J.; Wang, H.-J. Identifying Immature and Mature Pomelo Fruits in Trees by Elliptical Model Fitting in the Cr-Cb Color Space. *Precis. Agric.* **2019**, *20*, 138–156. [[CrossRef](#)]
95. Chen, J.; Lian, Y.; Zou, R.; Zhang, S.; Ning, X.; Han, M. Real-Time Grain Breakage Sensing for Rice Combine Harvesters Using Machine Vision Technology. *Int. J. Agric. Biol. Eng.* **2020**, *13*, 194–199. [[CrossRef](#)]
96. Liang, Y.; Lin, H.; Kang, W.; Shao, X.; Cai, J.; Li, H.; Chen, Q. Application of Colorimetric Sensor Array Coupled with Machine-learning Approaches for the Discrimination of Grains Based on Freshness. *J. Sci. Food Agric.* **2023**, *103*, 6790–6799. [[CrossRef](#)]
97. Fan, P.; Lang, G.; Guo, P.; Liu, Z.; Yang, F.; Yan, B.; Lei, X. Multi-Feature Patch-Based Segmentation Technique in the Gray-Centered RGB Color Space for Improved Apple Target Recognition. *Agriculture* **2021**, *11*, 273. [[CrossRef](#)]
98. Jiao, Y.; Luo, R.; Li, Q.; Deng, X.; Yin, X.; Ruan, C.; Jia, W. Detection and Localization of Overlapped Fruits Application in an Apple Harvesting Robot. *Electronics* **2020**, *9*, 1023. [[CrossRef](#)]
99. Xu, Y.; Imou, K.; Kaizu, Y.; Saga, K. Two-Stage Approach for Detecting Slightly Overlapping Strawberries Using HOG Descriptor. *Biosyst. Eng.* **2013**, *115*, 144–153. [[CrossRef](#)]
100. Feng, J.; Zeng, L.; He, L. Apple Fruit Recognition Algorithm Based on Multi-Spectral Dynamic Image Analysis. *Sensors* **2019**, *19*, 949. [[CrossRef](#)]
101. Dhakshina Kumar, S.; Esakkirajan, S.; Bama, S.; Keerthiveena, B. A Microcontroller Based Machine Vision Approach for Tomato Grading and Sorting Using SVM Classifier. *Microprocess. Microsyst.* **2020**, *76*, 103090. [[CrossRef](#)]
102. Peng, Y.; Zhao, S.; Liu, J. Fused Deep Features-Based Grape Varieties Identification Using Support Vector Machine. *Agriculture* **2021**, *11*, 869. [[CrossRef](#)]
103. Ling, X.; Zhao, Y.; Gong, L.; Liu, C.; Wang, T. Dual-Arm Cooperation and Implementing for Robotic Harvesting Tomato Using Binocular Vision. *Rob. Auton. Syst.* **2019**, *114*, 134–143. [[CrossRef](#)]
104. Reyes, J.F.; Contreras, E.; Correa, C.; Melin, P. Image Analysis of Real-Time Classification of Cherry Fruit from Colour Features. *J. Agric. Eng.* **2021**, *52*, 1160. [[CrossRef](#)]
105. Zhou, X.; Zhao, C.; Sun, J.; Cao, Y.; Yao, K.; Xu, M. A Deep Learning Method for Predicting Lead Content in Oilseed Rape Leaves Using Fluorescence Hyperspectral Imaging. *Food Chem.* **2023**, *409*, 135251. [[CrossRef](#)]

106. Xue, Y.; Jiang, H. Monitoring of Chlorpyrifos Residues in Corn Oil Based on Raman Spectral Deep-Learning Model. *Foods* **2023**, *12*, 2402. [\[CrossRef\]](#)
107. Huang, Y.; Li, Z.; Bian, Z.; Jin, H.; Zheng, G.; Hu, D.; Sun, Y.; Fan, C.; Xie, W.; Fang, H. Overview of Deep Learning and Nondestructive Detection Technology for Quality Assessment of Tomatoes. *Foods* **2025**, *14*, 286. [\[CrossRef\]](#)
108. Qiu, D.; Guo, T.; Yu, S.; Liu, W.; Li, L.; Sun, Z.; Peng, H.; Hu, D. Classification of Apple Color and Deformity Using Machine Vision Combined with CNN. *Agriculture* **2024**, *14*, 978. [\[CrossRef\]](#)
109. Pan, Z.; Gu, J.; Wang, W.; Fang, X.; Xia, Z.; Wang, Q.; Wang, M. Picking Point Identification and Localization Method Based on Swin-Transformer for High-Quality Tea. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 102262. [\[CrossRef\]](#)
110. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentations. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
111. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 6 January 2015.
112. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors* **2016**, *16*, 1222. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Wang, P.; Niu, T.; He, D. Tomato Young Fruits Detection Method under near Color Background Based on Improved Faster R-Cnn with Attention Mechanism. *Agriculture* **2021**, *11*, 1059. [\[CrossRef\]](#)
114. Wan, S.; Goudos, S. Faster R-CNN for Multi-Class Fruit Detection Using a Robotic Vision System. *Comput. Netw.* **2020**, *168*, 107036. [\[CrossRef\]](#)
115. Mai, X.; Zhang, H.; Jia, X.; Meng, M.Q.H. Faster R-CNN with Classifier Fusion for Automatic Detection of Small Fruits. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 1555–1569. [\[CrossRef\]](#)
116. Apolo-Apolo, O.E.; Martínez-Guanter, J.; Egea, G.; Raja, P.; Pérez-Ruiz, M. Deep Learning Techniques for Estimation of the Yield and Size of Citrus Fruits Using a UAV. *Eur. J. Agron.* **2020**, *115*, 126030. [\[CrossRef\]](#)
117. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [\[CrossRef\]](#)
118. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit Detection for Strawberry Harvesting Robot in Non-Structural Environment Based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [\[CrossRef\]](#)
119. Wang, D.; He, D. Fusion of Mask RCNN and Attention Mechanism for Instance Segmentation of Apples under Complex Background. *Comput. Electron. Agric.* **2022**, *196*, 106864. [\[CrossRef\]](#)
120. Lv, J.; Xu, H.; Xu, L.; Gu, Y.; Rong, H.; Zou, L. An Image Rendering-Based Identification Method for Apples with Different Growth Forms. *Comput. Electron. Agric.* **2023**, *211*, 108040. [\[CrossRef\]](#)
121. Wang, Y.; Lv, J.; Xu, L.; Gu, Y.; Zou, L.; Ma, Z. A Segmentation Method for Waxberry Image under Orchard Environment. *Sci. Hortic.* **2020**, *266*, 109309. [\[CrossRef\]](#)
122. Pei, H.; Sun, Y.; Huang, H.; Zhang, W.; Sheng, J.; Zhang, Z. Weed Detection in Maize Fields by UAV Images Based on Crop Row Preprocessing and Improved YOLOv4. *Agriculture* **2022**, *12*, 975. [\[CrossRef\]](#)
123. Magalhães, S.A.; Castro, L.; Moreira, G.; Dos Santos, F.N.; Cunha, M.; Dias, J.; Moreira, A.P. Evaluating the Single-Shot Multibox Detector and Yolo Deep Learning Models for the Detection of Tomatoes in a Greenhouse. *Sensors* **2021**, *21*, 3569. [\[CrossRef\]](#)
124. Xia, Z.; Zhao, Y.; Gu, J.; Wang, W.; Huang, Z.; Gao, Y.; Sun, P. The end-to-end chip surface defect segmentation method based on the diffusion model and attention mechanism. *Eng. Appl. Artif. Intell.* **2025**, *155*, 111131. [\[CrossRef\]](#)
125. Zhu, X.; Chen, F.; Zhang, X.; Zheng, Y.; Peng, X.; Chen, C. Detection the Maturity of Multi-Cultivar Olive Fruit in Orchard Environments Based on Olive-EfficientDet. *Sci. Hortic.* **2024**, *324*, 112607. [\[CrossRef\]](#)
126. Kirk, R.; Cielniak, G.; Mangan, M. L*a*b*Fruits: A Rapid and Robust Outdoor Fruit Detection System Combining Bio-Inspired Features with One-Stage Deep Learning Networks. *Sensors* **2020**, *20*, 275. [\[CrossRef\]](#)
127. Kang, S.; Hu, Z.; Liu, L.; Zhang, K.; Cao, Z. Object Detection YOLO Algorithms and Their Industrial Applications: Overview and Comparative Analysis. *Electronics* **2025**, *14*, 1104. [\[CrossRef\]](#)
128. Joseph, R.; Santosh, D.; Ross, G.; Ali, F. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NE, USA, 26 June–1 July 2016; pp. 779–788.
129. Junos, M.H.; Mohd Khairuddin, A.S.; Thannirmalai, S.; Dahari, M. Automatic Detection of Oil Palm Fruits from UAV Images Using an Improved YOLO Model. *Vis. Comput.* **2022**, *38*, 2341–2355. [\[CrossRef\]](#)
130. Zhang, Z.; Lu, Y.; Zhao, Y.; Pan, Q.; Jin, K.; Xu, G.; Hu, Y. TS-YOLO: An All-Day and Lightweight Tea Canopy Shoots Detection Model. *Agronomy* **2023**, *13*, 1411. [\[CrossRef\]](#)
131. Zhang, Q.; Chen, Q.; Xu, W.; Xu, L.; Lu, E. Prediction of Feed Quantity for Wheat Combine Harvester Based on Improved YOLOv5s and Weight of Single Wheat Plant without Stubble. *Agriculture* **2024**, *14*, 1251. [\[CrossRef\]](#)
132. Fu, L.; Feng, Y.; Wu, J.; Liu, Z.; Gao, F.; Majeed, Y.; Al-Mallahi, A.; Zhang, Q.; Li, R.; Cui, Y. Fast and Accurate Detection of Kiwifruit in Orchard Using Improved YOLOv3-Tiny Model. *Precis. Agric.* **2021**, *22*, 754–776. [\[CrossRef\]](#)

133. Ji, W.; Gao, X.; Xu, B.; Pan, Y.; Zhang, Z.; Zhao, D. Apple Target Recognition Method in Complex Environment Based on Improved YOLOv4. *J. Food Process Eng.* **2021**, *44*, e13866. [\[CrossRef\]](#)
134. Ji, W.; Pan, Y.; Xu, B.; Wang, J. A Real-Time Apple Targets Detection Method for Picking Robot Based on ShufflenetV2-YOLOX. *Agriculture* **2022**, *12*, 856. [\[CrossRef\]](#)
135. Zhang, F.; Chen, Z.; Ali, S.; Yang, N.; Fu, S.; Zhang, Y. Multi-Class Detection of Cherry Tomatoes Using Improved YOLOv4-Tiny. *Int. J. Agric. Biol. Eng.* **2023**, *16*, 225–231. [\[CrossRef\]](#)
136. Hu, T.; Wang, W.; Gu, J.; Xia, Z.; Zhang, J.; Wang, B. Research on Apple Object Detection and Localization Method Based on Improved YOLOX and RGB-D Images. *Agronomy* **2023**, *13*, 1816. [\[CrossRef\]](#)
137. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [\[CrossRef\]](#)
138. Kaukab, S.; Ghodki, B.M.; Ray, H.; Kalnar, Y.B.; Narsaiah, K.; Brar, J.S. Improving Real-Time Apple Fruit Detection: Multi-Modal Data and Depth Fusion with Non-Targeted Background Removal. *Ecol. Inform.* **2024**, *82*, 102691. [\[CrossRef\]](#)
139. Lawal, O.M. YOLOv5-LiNet: A Lightweight Network for Fruits Instance Segmentation. *PLoS ONE* **2023**, *18*, e0282297. [\[CrossRef\]](#)
140. Tao, T.; Wei, X. STBNA-YOLOv5: An Improved YOLOv5 Network for Weed Detection in Rapeseed Field. *Agriculture* **2024**, *15*, 22. [\[CrossRef\]](#)
141. Xu, B.; Cui, X.; Ji, W.; Yuan, H.; Wang, J. Apple Grading Method Design and Implementation for Automatic Grader Based on Improved YOLOv5. *Agriculture* **2023**, *13*, 124. [\[CrossRef\]](#)
142. Zhang, T.; Zhou, J.; Liu, W.; Yue, R.; Yao, M.; Shi, J.; Hu, J. Seedling-YOLO: High-Efficiency Target Detection Algorithm for Field Broccoli Seedling Transplanting Quality Based on YOLOv7-Tiny. *Agronomy* **2024**, *14*, 931. [\[CrossRef\]](#)
143. Gu, B.; Wen, C.; Liu, X.; Hou, Y.; Hu, Y.; Su, H. Improved YOLOv7-Tiny Complex Environment Citrus Detection Based on Lightweighting. *Agronomy* **2023**, *13*, 2667. [\[CrossRef\]](#)
144. Wang, W.; Xi, Y.; Gu, J.; Yang, Q.; Pan, Z.; Zhang, X.; Xu, G.; Zhou, M. YOLOv8-TEA: Recognition Method of Tender Shoots of Tea Based on Instance Segmentation Algorithm. *Agronomy* **2025**, *15*, 1318. [\[CrossRef\]](#)
145. Zhang, G.; Yang, X.; Lv, D.; Zhao, Y.; Liu, P. YOLOv8n-CSD: A Lightweight Detection Method for Nectarines in Complex Environments. *Agronomy* **2024**, *14*, 2427. [\[CrossRef\]](#)
146. Ma, B.; Hua, Z.; Wen, Y.; Deng, H.; Zhao, Y.; Pu, L.; Song, H. Using an Improved Lightweight YOLOv8 Model for Real-Time Detection of Multi-Stage Apple Fruit in Complex Orchard Environments. *Artif. Intell. Agric.* **2024**, *11*, 70–82. [\[CrossRef\]](#)
147. Ma, J.; Zhao, Y.; Fan, W.; Liu, J. An Improved YOLOv8 Model for Lotus Seedpod Instance Segmentation in the Lotus Pond Environment. *Agronomy* **2024**, *14*, 1325. [\[CrossRef\]](#)
148. Xie, H.; Zhang, Z.; Zhang, K.; Yang, L.; Zhang, D.; Yu, Y. Research on the Visual Location Method for Strawberry Picking Points under Complex Conditions Based on Composite Models. *J. Sci. Food Agric.* **2024**, *104*, 8566–8579. [\[CrossRef\]](#)
149. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single Shot Multibox Detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
150. Wang, Y.; Xing, Z.; Ma, L.; Qu, A.; Xue, J. Object Detection Algorithm for Lingwu Long Jujubes Based on the Improved SSD. *Agriculture* **2022**, *12*, 1456. [\[CrossRef\]](#)
151. Agarwal, D.; Bhargava, A. On-Tree Fruit Detection System Using Darknet-19 Based SSD Network. *J. Food Meas. Charact.* **2024**, *18*, 7067–7076. [\[CrossRef\]](#)
152. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 213–229.
153. Yang, Y.; Wang, X.; Liu, Z.; Huang, M.; Sun, S.; Zhu, Q. Detection of Multi-Size Peach in Orchard Using RGB-D Camera Combined with an Improved DETection Transformer Model. *Intell. Data Anal.* **2023**, *27*, 1539–1554. [\[CrossRef\]](#)
154. Huang, Z.; Zhang, X.; Wang, H.; Wei, H.; Zhang, Y.; Zhou, G. Pear Fruit Detection Model in Natural Environment Based on Lightweight Transformer Architecture. *Agriculture* **2024**, *15*, 24. [\[CrossRef\]](#)
155. Ji, W.; Zhai, K.; Xu, B.; Wu, J. Green Apple Detection Method Based on Multidimensional Feature Extraction Network Model and Transformer Module. *J. Food Prot.* **2025**, *88*, 100397. [\[CrossRef\]](#)
156. Zhang, J.; Gu, J.; Hu, T.; Wang, B.; Xia, Z. An Image Segmentation and Point Cloud Registration Combined Scheme for Sensing of Obscured Tree Branches. *Comput. Electron. Agric.* **2024**, *221*, 108960. [\[CrossRef\]](#)
157. Chen, Z.; Ting, D.; Newbury, R.; Chen, C. Semantic Segmentation for Partially Occluded Apple Trees Based on Deep Learning. *Comput. Electron. Agric.* **2021**, *181*, 105952. [\[CrossRef\]](#)
158. Li, J.; Tang, Y.; Zou, X.; Lin, G.; Wang, H. Detection of Fruit-Bearing Branches and Localization of Litchi Clusters for Vision-Based Harvesting Robots. *IEEE Access* **2020**, *8*, 117746–117758. [\[CrossRef\]](#)
159. Song, Z.; Zhou, Z.; Wang, W.; Gao, F.; Fu, L.; Li, R.; Cui, Y. Canopy Segmentation and Wire Reconstruction for Kiwifruit Robotic Harvesting. *Comput. Electron. Agric.* **2021**, *181*, 105933. [\[CrossRef\]](#)
160. Lin, G.; Tang, Y.; Zou, X.; Wang, C. Three-Dimensional Reconstruction of Guava Fruits and Branches Using Instance Segmentation and Geometry Analysis. *Comput. Electron. Agric.* **2021**, *184*, 106107. [\[CrossRef\]](#)

161. Yang, C.H.; Xiong, L.Y.; Wang, Z.; Wang, Y.; Shi, G.; Kuremot, T.; Zhao, W.H.; Yang, Y. Integrated Detection of Citrus Fruits and Branches Using a Convolutional Neural Network. *Comput. Electron. Agric.* **2020**, *174*, 105469. [[CrossRef](#)]
162. Wan, H.; Fan, Z.; Yu, X.; Kang, M.; Wang, P.; Zeng, X. A Real-Time Branch Detection and Reconstruction Mechanism for Harvesting Robot via Convolutional Neural Network and Image Segmentation. *Comput. Electron. Agric.* **2022**, *192*, 106609. [[CrossRef](#)]
163. Zheng, Z.; Hu, Y.; Guo, T.; Qiao, Y.; He, Y.; Zhang, Y.; Huang, Y. AGHRNet: An Attention Ghost-HRNet for Confirmation of Catch-and-shake Locations in Jujube Fruits Vibration Harvesting. *Comput. Electron. Agric.* **2023**, *210*, 107921. [[CrossRef](#)]
164. Zheng, Z.; Liu, Y.; Dong, J.; Zhao, P.; Qiao, Y.; Sun, S.; Huang, Y. A Novel Jujube Tree Trunk and Branch Salient Object Detection Method for Catch-and-Shake Robotic Visual Perception. *Expert. Syst. Appl.* **2024**, *251*, 124022. [[CrossRef](#)]
165. Kok, E.; Wang, X.; Chen, C. Obscured Tree Branches Segmentation and 3D Reconstruction Using Deep Learning and Geometrical Constraints. *Comput. Electron. Agric.* **2023**, *210*, 107884. [[CrossRef](#)]
166. Digumarti, S.T.; Nieto, J.; Cadena, C.; Siegwart, R.; Beardsley, P. Automatic Segmentation of Tree Structure From Point Cloud Data. *IEEE Robot. Autom. Lett.* **2018**, *3*, 3043–3050. [[CrossRef](#)]
167. Ma, B.; Du, J.; Wang, L.; Jiang, H.; Zhou, M. Automatic Branch Detection of Jujube Trees Based on 3D Reconstruction for Dormant Pruning Using the Deep Learning-Based Method. *Comput. Electron. Agric.* **2021**, *190*, 106484. [[CrossRef](#)]
168. Westling, F.; Underwood, J.; Bryson, M. Graph-Based Methods for Analyzing Orchard Tree Structure Using Noisy Point Cloud Data. *Comput. Electron. Agric.* **2021**, *187*, 106270. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.